(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36645 A2

(51) International Patent Classification[7]: C12N 15/54, 15/56, 9/50, 9/12, C07K 16/40, C12Q 1/68, G01N 33/53, 33/68, A61K 38/17

(21) International Application Number: PCT/US00/31744

(22) International Filing Date:
17 November 2000 (17.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/165,986	17 November 1999 (17.11.1999)	US
60/181,347	9 February 2000 (09.02.2000)	US
60/194,195	3 April 2000 (03.04.2000)	US
60/194,839	5 April 2000 (05.04.2000)	US
60/195,637	7 April 2000 (07.04.2000)	US
60/197,080	13 April 2000 (13.04.2000)	US
60/232,677	15 September 2000 (15.09.2000)	US
Not furnished	16 November 2000 (16.11.2000)	US

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 11th floor, 555 Long Wharf Drive, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **QUINN, Kerry, E.** [US/US]; 51 Greenfield Drive, Torrington, CT 06790 (US). **SPYTEK, Kimberley, Ann** [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). **MAJUMDER, Kumud** [US/US]; 140 Silver Hill Lane, Stamford, CT 06905 (US). **VERNET, Corine** [FR/US]; 4830 NW 43rd Street P#253, Gainesville, FL 32060 (US). **BURGESS, Catherine** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). **FERNANDES, Elma** [IN/US]; 77 Florence Road #2B, Branford, CT 06405 (US). **TAUPIER, Raymond, Jr.** [US/US]; 47 Holmes Street, East Haven, CT 06512 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **HERRMANN, John, L.** [US/US]; 78 Barnshed Lane, Guilford, CT 06437 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/36645 A2

(54) Title: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: The present invention provides novel isolated SERX polynucleotides and polypeptides encoded by the SERX polynucleotides. Also provided are the antibodies that immunospecifically bind to a SERX polypeptide or any derivative, variant, mutant or fragment of the SERX polypeptide, polynucleotide or antibody. The invention additionally provides methods in which the SERX polypeptide, polynucleotide and antibody are utilized in the detection and treatment of a broad range of pathological states, as well as to other uses.

NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom. More specifically, the invention relates to nucleic acids encoding membrane bound and secreted polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based, in part, upon the discovery of a novel polynucleotide sequences encoding novel polypeptides.

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule that includes the sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide that includes the amino acid sequences of SEQ ID NO: 2, 6, 9, 11, or 15. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes a SERX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified SERX polypeptide, *e.g.*, any of the SERX polypeptides encoded by a SERX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a SERX polypeptide and a pharmaceutically acceptable carrier or diluent.

In still a further aspect, the invention provides an antibody that binds specifically to a SERX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including SERX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing a SERX polypeptide by providing a cell containing a SERX nucleic acid, *e.g.*, a vector that includes a SERX nucleic acid, and culturing the cell under conditions sufficient to express the SERX polypeptide encoded by the nucleic acid. The expressed SERX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous SERX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying a SERX polypeptide or nucleic acid in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a SERX polypeptide by contacting a SERX polypeptide with a compound and determining whether the SERX polypeptide activity is modified.

The invention is also directed to compounds that modulate SERX polypeptide activity identified by contacting a SERX polypeptide with the compound and determining whether the compound modifies activity of the SERX polypeptide, binds to the SERX polypeptide, or binds to a nucleic acid molecule encoding a SERX polypeptide.

In another aspect, the invention provides a method of determining the presence of or predisposition of a SERX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of SERX polypeptide in the subject sample. The amount of SERX polypeptide in the subject sample is then compared to the amount of SERX polypeptide in a control sample. For example, an alteration in the amount of SERX polypeptide in the subject protein sample relative to the amount of SERX polypeptide

in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the SERX is detected using a SERX antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of a SERX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the SERX nucleic acid in the subject nucleic acid sample. The amount of SERX nucleic acid sample in the subject nucleic acid is then compared to the amount of a SERX nucleic acid in a control sample. An alteration in the amount of SERX nucleic acid in the sample relative to the amount of SERX in the control sample indicates the subject has a tissue proliferation-associated disorder.

In a still further aspect, the invention provides a method of treating or preventing or delaying a SERX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a SERX nucleic acid, a SERX polypeptide, or a SERX antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences and their polypeptides. The sequences are collectively referred to as "SERX nucleic acids" or "SERX polynucleotides" and the corresponding encoded polypeptides are referred to as "SERX polypeptides" or "SERX proteins." Unless indicated otherwise, "SERX" is meant to refer to any of the novel sequences disclosed herein. Table 1 provides a summary of the SERX nucleic acids and their encoded polypeptides.

Table 1 Sequences and Corresponding SEQ ID Numbers

SERX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1	57660562	1	2	Serine/Threonine Kinase
2	AC010431_A	5	6	Serine/Threonine Kinase
3	AC010431_da1	7	6	Serine/Threonine Kinase
4	24111358_EXT1	8	9	Serine/Threonine Kinase
5	GM_10221687_A	10	11	Serine Protease
6	12996895_1	14	15	Serine Protease
7	12996895.0.1	16	NA	Serine Protease
8	249A	17	NA	Serine Protease

SERX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various SERX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, SERX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the SERX polypeptides belong.

For example, SER1, SER2, SER3, and SER4 are homologous to members of the serine/threonine kinase family of proteins. Thus, the SER1-4 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in disorders associated with, *e.g.*, signal transduction pathways, cell proliferation, growth, and spermiogenesis.

Protein phosphorylation is a fundamental process for the regulation of cellular functions. The coordinated action of both protein kinases and phosphatases controls the levels of phosphorylation and, hence, the activity of specific target proteins. One of the predominant roles of protein phosphorylation is in signal transduction, where extracellular signals are amplified and propagated by a cascade of protein phosphorylation and dephosphorylation events. Two of the best characterized signal transduction pathways involve the cAMP-dependent protein kinase and protein kinase C (PKC). Each pathway uses a different second-messenger molecule to activate the protein kinase, which, in turn, phosphorylates specific target molecules.

Most of the effects of cAMP in the eukaryotic cell are mediated through the phosphorylation of target proteins on serine or threonine residues by the cAMP-dependent protein kinase. The inactive cAMP-dependent protein kinase is a tetramer composed of 2 regulatory and 2 catalytic subunits. The cooperative binding of 4 molecules of cAMP dissociates the enzyme in a regulatory subunit dimer and 2 free active catalytic subunits. In the human, 4 different regulatory subunits (PRKAR1A, OMIM Acc.188830; PRKAR1B, OMIM Acc.176911; PRKAR2A, OMIM Acc.176910; and PRKAR2B, OMIM Acc.176912) and 3 catalytic subunits (PRKACA; PRKACB, OMIM Acc.176892; and PRKACG OMIM Acc.176893) have been identified. The OMIM accession numbers refer to entries in the Online Mendelian Genetics database, a catalog of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information. See, URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM.

Extensive comparisons of kinase sequences defined a common catalytic domain, ranging from 250 to 300 amino acids. This domain contains key amino acids conserved between kinases and are thought to play an essential role in catalysis. Jones et al. (Cell Regul. 2:1001-1009 (1991) isolated a partial cDNA that encodes a protein kinase they termed rac (related to the A and C kinases). DNA sequencing identified an open reading frame of 1,440 bp encoding a protein of 480 amino acids. In an in vitro translation system that used RNA transcribed from cloned cDNAs, they demonstrated the synthesis of a protein of corresponding size. For example, a testis-specific novel Ser/Thr kinase-1 (TSK-1) for the gene reveals highest homology to the human gene encoding rac protein kinase-beta and the group of yeast Ser/Thr

kinases encoded by SNF-1,nim-1, KIN-1 and KIN-2. Other serine/threonine protein kinases contain consensus sequences characteristic of a protein kinase catalytic domain and showed 73 and 68% similarity to protein kinase C and cAMP-dependent protein kinase, respectively. Other serine/threonine protein kinases include RAF1 and PIM1.

SER5, SER6, SER7, and SER8 are homologous to members of the serine protease family of proteins. Thus, the SER5-8 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications in *e.g.*, blood clotting disorders, cell growth and proliferative disorders.

The SERX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance SERX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, cell growth and differentiation protein processing.

Additional utilities for SERX nucleic acids and polypeptides according to the invention are disclosed herein.

SER1

A SER1 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the serine/threonine kinase family of proteins. A SER1 nucleic acid and its encoded polypeptide includes the sequences shown in Table 2 and Table 3, respectively. The disclosed nucleic acid (SEQ ID NO:1) is 823 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 142-144 and ends with a TGA stop codon at nucleotides 772-774. The representative ORF includes a 210 amino acid polypeptide (SEQ ID NO:2).

Table 2. Nucleotide sequence including the sequence encoding the serine/ threonine kinase-like protein of the invention. SER1

CACAGCGGCCAATGTCTGGCAGTGGGCACATGGGGGTCGGGGGGTGTAGGTGCCAAGCGCCATGGCTTAGACCCGAGA
TTGGAGTCCGGCCGCCCCCCGACAGCAGCCGCCTCCTGCCCCCCGTGCGCCCTAGGCGCCACC**ATG**TCGGGAGACAAA
CTTCTGAGCGAACTCGGTTATAAGCTGGGCCGCACAATTGGAGAGGGCAGCTACTCCAAGGTGAAGGTGGCCACATCC
AAGAAGTACAAGGGTACCGTGGCCATCAAGGTGGTGGACCGGCGGCGAGCGCCCCCGGACTTCGTCAACAAGTTCCTG
CCGCGAGAGCTGTCCATCCTGCGGGGCGTGCGACACCCGCACATCGTGCACGTCTTCGAGTTCATCGAGGTGTGCAAC
GGGAAACTGTACATCGTGATGGAAGCGGCCGCCACCGACCTGCTGCAAGCCGTGCAGCGCAACGGGCGCATCCCCGGA
GTTCAGGCGCGCGACCTCTTTGCGCAGATCGCCGGCGCCGTGCGCTACCTGCACGATCATCACCTGGTGCACCGCGAC
CTCAAGTGCGAAAACGTGCTGCTGAGCCCGGACGAGCGCCGCGTCAAGCTCACCGACTTCGGCTTCGGCCGCCAGGCC
CATGGCTACCCAGACCTGAGCACCACCTACTGCGGCTCAGCCGTACGCGTCACCCGAGTCATGCATTTCTTGAGCACC
TACTGTCTGCCAGGCCCCAGAGCTCATGGCGAAGAGACTTGGGCCCATCCCTGCCGAAAACGAGACAATT**GA**AAAGTC
AAGTAAAATAAAAGAATGACATGGAAATAAAAAAAAAAAAAAA (SEQ ID NO:1).

Table 3. Protein sequence for SER1.

MSGDKLLSELGYKLGRTI**GEGSYSKVKVATSKKYKGTVAIK**VVDRRRAPPDFVNKFLPRELSILRGVRHPHIVHVFEF
IEVCNGKLYIVMEAAATDLLQAVQRNGRIPGVQARDLFAQIAGAVRYLHDHH***LVHRDLKCENVLLS***PDERRVKLTDFG
FGRQAHGYPDLSTTYCGSAVRVTRVMHFLSTYCLPGPRAHGEETWAHPCRKRDN (SEQ ID NO:2)

In a search of the public sequence databases, leading to the identification of the serine/threonine kinase-like protein SER1, it was found that public ESTs cover all but a 69 bp region in the center of the gene (at bases 422-491 of SeqCallingTM assembly 57660562). Therefore, the full clone for SER1 is only able to be completed using fragments obtained using CuraGen's SeqCallingTM sequencing procedure. SeqCalling is a differential expression and sequencing procedure that normalizes mRNA species in a sample, and is disclosed in U. S. Ser. No. 09/417,386, filed Oct. 13, 1999, incorporated herein by reference in its entirety.

A hydrophobicity plot for SER1 suggests that the gene encodes a protein not having a signal peptide at the N-terminus. The results suggest that the serine/threonine kinase-like protein of the invention may be localized in one or another intracellular organelle, or the nucleus. The presence of two regions of positive hydrophobicity together with other analyses suggests that the serine-threonine kinase-like protein described herein is an integral membrane protein.

Among tissues from which an mRNA encoding the serine/threonine kinase-like protein of the invention may be obtained are testis, fetal lung, B-cells, kidney, lung (such as lung tumor), prostate, lymphocyte, brain, spleen, and pancreas.

BLASTX comparisons with known proteins showed that the highest percent positives, 64%, were found for the genotypes embodying cardiovascular abnormalities such as aortic arch abnormalities involved in DiGeorge syndrome (DGS; Lindsay et al. Nature 401:379-83 (1999) and velo-cardiofacial syndrome (VCFS; Kimber et al. Hum Mol Genet 8:2229-2237 (1999)).

The full amino acid sequence of the protein of the invention was found to have 83 of 179 amino acid residues (46 %) identical to, and 115 of 179 residues (64 %) positive with, the 365 amino acid residue serine-threonine kinase from the mouse (ptnr: SPTREMBL-ACC:P97416):

•Sptrembl-Acc:P97416 Serine/threonine kinase - *Mus musculus*, 365 aa.

Score = 398 (140.1 Bits), Expect = 3.0e-36, P = 3.0e-36; Identities = 83/179 (46%); Positives = 115/179 (64%), Frame = +1

•Sptrembl-Acc:Q61241 Testis-specific serine/threonine kinase - *Mus musculus*, 364 aa.

Score = 382 (134.5 Bits), Expect = 2.0e-34, P = 2.0e-34; Identities = 81/179 (45%), Positives = 113/179 (63%), Frame = +1

•Sptrembl-Acc:O54863 Testis specific serine/threonine kinase 2- *Mus musculus*, 357 aa.

Score = 380 (133.8 Bits), Expect = 3.3e-34, P = 3.3e-34; Identities = 80/179 (44%); Positives = 112/179 (62%), Frame = +1.

A search of the PROSITE database of protein families and domains confirmed that a SER1 polypeptide is a member of the Eukaryotic protein kinase family. Eukaryotic protein kinases are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. Hanks & Hunter FASEB J. 9:576-596(1995); Hanks et al., Science 241:42-52(1988). There are a number of conserved regions in the catalytic domain of protein kinases. One region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a

conserved aspartic acid residue which is important for the catalytic activity of the enzyme; there are two signature patterns for this region: one specific for serine/ threonine kinases and the other for tyrosine kinases. Knighton et al., Science 253:407-414(1991).

The ATP binding pattern is: [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x- [GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K (SEQ ID NO:3).

The majority of known protein kinases belong to the class detected by this pattern. However, viral kinases are quite divergent in this region and may be completely missed by this pattern. This pattern is found in amino acids 18-41 of SEQ ID NO:2 (shown in bold).

The second consensus pattern, located in the central part of the catalytic domain, is: [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3) (SEQ ID NO:4), where D is an active site residue. This pattern is indicated by the bold, italic residues 131-143 of SEQ ID NO:2.

The SER1 polypeptide sequence contains both Eukaryotic protein kinase signature sequences as defined by the PROSITE database (illustrated by bold and bold, italics in SEQ ID NO:2 of Table 3). The protein kinase ATP binding signature sequence is located at amino acid residues 18-41 of SEQ ID NO:2 (bold). The serine/threonine protein kinase active site signature sequence is located at amino acid residues 131-143 of SEQ ID NO:2 (bold, italic). The PROSITE database consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family a new sequence belongs.

Additionally, a search of the Pfam protein families database of alignments shows that SEQ ID NO:2 belongs to the Eukaryotic protein kinase family, identified by Accession Number PF00069.

Furthermore, a search of the NCBI CD-Search database demonstrated that SER1 shows significant similarity to three kinase domains, as shown in Table 4. The NCBI CD-Search is a search of a conserved domain database. Because proteins often contain several modules or domains, each with a distinct evolutionary origin and function, the CD-Search service may be used to identify the conserved domains present in a protein sequence. Computational biologists define conserved domains based on recurring sequence patterns or motifs. CDD currently contains domains derived from two popular collections, Smart and Pfam, plus

contributions from NCBI. To identify conserved domains in a protein sequence, the CD-Search service employs the reverse position-specific BLAST algorithm. The query sequence is compared to a position-specific score matrix prepared from the underlying conserved domain alignment. Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997).

Table 4 CDD Domain Homology

Sequences producing significant alignments	Score (bits)	E value	Row
Gnl|Smart|S_TKc: Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily	148	2e-37	1
Gnl|Pfam|pfam00069: pkinase, Eukaryotic protein kinase domain...	133	8e-33	2
Gnl|Smart|TyrKc: Tyrosine kinase, catalytic domain; Phosphotransferases, Tyrosine-specific kinase subfamily.	92.4	2e-20	3

This degree of homology between a SER1 polypeptide and the kinase polypeptide domains (both in terms of length and complexity) is very unlikely to have occurred by chance alone (e.g., the Expect (E) value in Table 4 less than $2e^{-37}$ by chance for the serine/threonine protein kinase catalytic domain).

The CD multiple Align for SER1 to the catalytic domain (row 1, Table4) is shown in Table 5a. The figure indicates that SER1 has homology to 'gi 5542282' (accession number P48479), a c-Jun N-terminal kinase implicated in neuronal apoptosis. Xie et al., Structure 6:983-991 (1998). Also shown in Table 5a is entry 'gi 1730584' (accession number P54739), a serine/threonine kinase from *Streptomyces coelicolor A3(2)*. Urabe & Ogawara Gene 153:99-104 (1995). Entry 'gi 125529' (accession number P27448) is a human putative serine/threonine protein kinase, P78. Finally, entry 'gi 6678167' (accession number NP_033462) is serine/threonine kinase 22B, which is associated with spermatogenesis. Kueng et al., J. Cell. Biol. 139:1851-1859 (1997). In the alignment figures, black outlined amino acids indicate identity (e.g., regions of conserved sequence that may be required to preserve structural or functional properties), whereas gray outline amino acids indicate conservative substitutions, (e.g., the residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function, e.g. L to V, I, or M); non-highlighted amino acid residues vary can likely be mutated to a much broader extent without altering structure or function.

The CD multiple Align for SER1 to the pkinase ATP binding domain (Row 2, Table 4) is shown in Table 5b. The figure indicates that SER1 has homology to 'gi 3318993,' the catalytic subunit of protein kinase Ck2 (casein kinase 2) from *Zea mays* involved in cell proliferation. Niefind et al., EMBO J 17:2451-2462 (1998). Also shown in Table 5b is entry 'gi 125690' (accession number P18653), a mouse ribosomal protein S6 kinase alpha 1, which is implicated in the activation of the mitogen-activated kinase cascade and belongs to the serine/threonine family of protein kinases. Alcorta et al., Mol. Cell. Biol. 9:3850-3859 (1989). Entry 'gi 1170662' (accession number P22987) is a yeast serine/threonine protein kinase, KIN1, which is important for growth polarity in *S. pombe*. Levin & Bishop, Proc. Natl. Acad, Sci., USA 87:8272-8276 (1990). Finally, entry 'gi 120621' (accession number P23647) is serine/threonine protein kinase from *Drosophila melanogaster*, which is a segment polarity protein important for correct patterning in the posterior part of each embryonic metamere. Preat et al., Nature 347:87-89 (1990).

Table 5a: CD Multiple Align- catalytic domain:

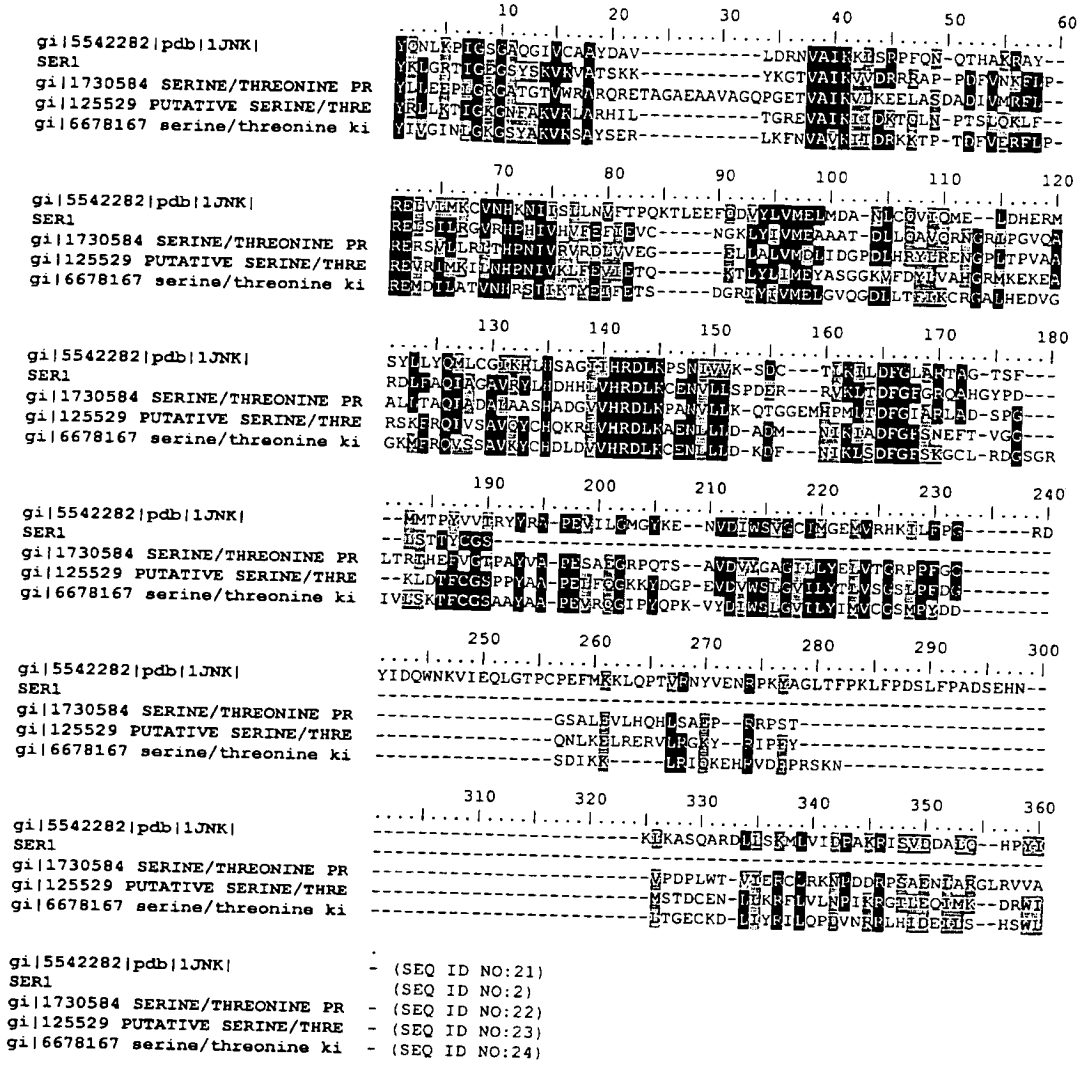
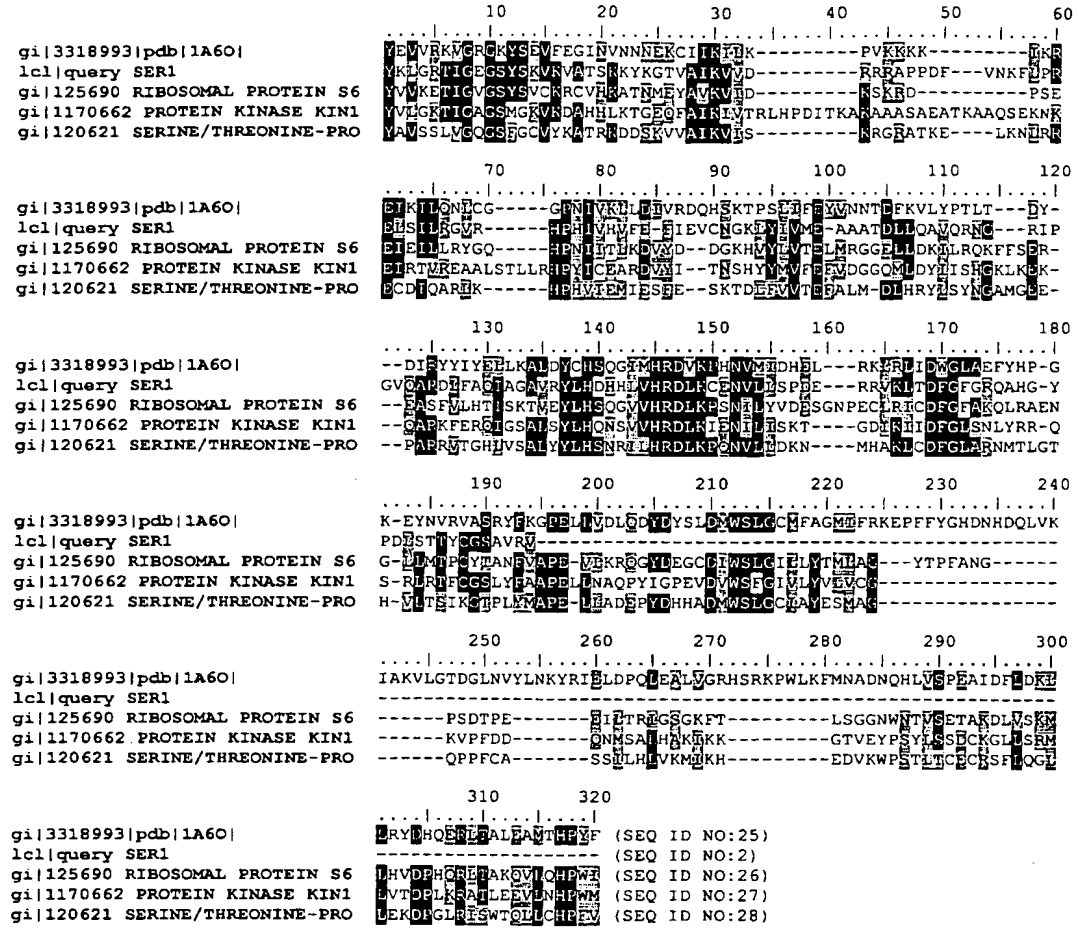


Table 5b: CD Multiple Align- pkinase domain:



The serine/threonine kinases comprise a family of structurally related proteins implicated in the regulation of a variety of diseases including those related to cell signal transduction disorders. This family of proteins is involved in developmental processes, and other significant physiological roles such as neuronal apoptosis, cell proliferation and spermiogenesis. Serine/threonine kinases are expressed in a number of different organisms and cell types, including for example, intestine, thymus, lung, fetal lung, lung tumor, B-cells, prostate, lymphocyte, brain, pancreas, spleen, bursa, and testis.

Based on its relatedness to the conserved serine/threonine kinase proteins, the SER1 protein is a novel member of the serine/threonine protein family. The discovery of molecules

related to serine/threonine kinases satisfies a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of serine/threonine kinase- like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in a variety of diseases and pathologies, including by way of nonlimiting example, those involving cardiovascular defects such as DGS syndrome and VCFS, hyperproliferative diseases, hypoproliferative diseases, as well as those involving issues relating to fertility.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in vascular, skeletal and cardiac muscle disorders. For example, a nucleic acid encoding the serine/threonine kinase-like protein may be useful in gene therapy, an inhibitory oligonucleotide directed to a mutant gene such as that involved in DGS or VCFS may block expression of the defective gene, and the serine/threonine kinase-like protein may be useful as complementation therapy when administered to a subject in need thereof. The novel nucleic acid encoding serine/threonine kinase-like protein, and the serine/threonine kinase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

A SER1 nucleic acid is useful for detecting specific cell-types. For example a SER1 nucleic acid according to the invention can be present in different levels in a lung tumor. Also, according to the invention the expression of a SER1 nucleic acid has utility in identifying developing and embryonic tissues from other tissue types.

SER2 & SER3

A SER2 sequence according to the invention is also a nucleic acid sequence encoding a polypeptide related to the serine/threonine family of proteins. A SER2 nucleic acid and its encoded polypeptide includes the sequences shown in Table 6. The disclosed nucleic acid (SEQ ID NO:5) is 1140 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 21-23 and ends with a TGA stop codon at nucleotides 1122-1124. The representative ORF includes a 367 amino acid polypeptide (SEQ ID NO:6) as shown in Table 7.

Table 6. Nucleotide sequence of SER2, the serine-threonine kinase -like protein of the invention.

CACTGGGCATTCCTGGCACC**ATG**GATGACGCTGCTGTCCTCAAGCGACGAGGCTACCTCCTGGGGATAAA
TTTAGGAGAGGGCTCCTATGCAAAAGTAAAATCTGCTTACTCTGAGCGCCTGAAGTTCAATGTGGCGATC
AAGATCATCGACCGCAAGAAGGCCCCCGCAGACTTCTTGGAGAAATTCCTTCCCCGGGAAATTGAGATTC
TGGCCATGTTAAACCACTGCTCCATCATTAAGACCTACGAGATCTTTGAGACATCACATGGCAAGGTCTA
CATCGTCATGGAGCTCGCGGTCCAGGGCGACCTCCTCGAGTTAATCAAAACCCGGGGAGCCCTGCATGAG
GACGAAGCTCGCAAGAAGTTCCACCAGCTTTCCTTGGCCATCAAGTACTGCCACGACCTGGACGTCGTCC
ACCGGGACCTCAAGTGTGACAACCTTCTCCTTGACAAGGACTTCAACATCAAGCTGTCCGACTTCAGCTT
CTCCAAGCGCTGCCTGCGGGATGACAGTGGTCGAATGGCATTAAGCAAGACCTTCTGTGGGTCACCAGCG
TATGCGGCCCCAGAGGTGCTGCAGGGCATTCCCTACCAGCCCAAGGTGTACGACATCTGGAGCCTAGGCG
TGATCCTCTACATCATGGTCTGCGGCTCCATGCCCTACGACGACTCCAACATCAAGAAGATGCTGCGTAT
CCAGAAGGAGCACCGCGTCAACTTCCCACGCTCCAAGCACCTGACAGGCGAGTGCAAGGACCTCATCTAC
CACATGCTGCAGCCCGACGTCAACCGGCGGCTCCACATCGACGAGATCCTCAGCCACTGCTGGATGCAGC
CCAAGGCACGGGGATCTCCCTCTGTGGCCATCAACAAGGAGGGGGAGAGTTCCCGGGGAACTGAACCCTT
GTGGACCCCCGAACCTGGCTCTGACAAGAAGTCTGCCACCAAGCTGGAGCCTGAGGGAGAGGCACAGCCC
CAGGCACAGCCTGAGACAAAACCCGAGGGGACAGCAATGCAAATGTCCAGGCAGTCGGAGATCCTGGGTT
TCCCCAGCAAGCCGTCGACTATGGAGACAGAGGAAGGGCCCCCCCAACAGCCTCCAGAGACGCGGGCCCA
G**TGA**GCTTCTTGCGGCCCAG (SEQ ID NO:5).

Table 7. Protein sequence encoded by SER2.

MDDAAVLKRRGYLLGINL**GEGSYAKVKSAYSERLKFNVAIK**IIDRKKAPADFLEKFLPREIEILAMLNHC
SIIKTYEIFETSHGKVYIVMELAVQGDLLELIKTRGALHEDEARKKFHQLSLAIKYCHDLD***VVHRDLKCD***
NLLLDKDFNIKLSDFSFSKRCLRDDSGRMALSKTFCGSPAYAAPEVLQGIPYQPKVYDIWSLGVILYIMV
CGSMPYDDSNIKKMLRIQKEHRVNFPRSKHLTGECKDLIYHMLQPDVNRRLHIDEILSHCWMQPKARGSP
SVAINKEGESSRGTEPLWTPEPGSDKKSATKLEPEGEAQPQAQPETKPEGTAMQMSRQSEILGFPSKPST
METEEGPPQQPPETRAQ (SEQ ID NO:6).

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of Acc. No. AC010431_A) has 986 of 1140 bases (86 %) identical to a M. musculus species Serine-Threonine Kinase mRNA (GENBANK-ID: U01840). The nucleic acid also has homology (approximately 88% identity) to mouse serine/threonine kinase 22A (spermiogenesis associated, reference NM_009435.1) and to mouse serine/threonine kinase (tsk-1) mRNA (88% identity, GenBank Accession. No. U01840). The encoded polypeptide

has 82% identity to serine/threonine kinase 22A (spermiogenesis associated, GenBank Accession AAA99535.1) and 67% identity to mouse serine/threonine kinase 22B (spermiogenesis associated, GenBank Accession AAC03367.1).

The target sequence identified as Accession Number AC010431_A (SEQ ID NO:5) was subjected to the exon linking process to confirm the sequence, as follows. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a library containing a wide range of cDNA species. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide a consensus sequence which is designated Accession Number AC010431_da1 (SEQ ID NO:7, SER3 as shown in Table 8). The coding sequence of clone AC010431_da1 was 100% identical to that provided in the sequence described as Acc. No. AC010431_A.

Table 8: Nucleic Acid Sequence for SER3

```
TTGGCACCATGGATGACGCTGCTGTCCTCAAGCGACGAGGCTACCTCCTGGGGATAAATTTAGGAGAGGG
CTCCTATGCAAAAGTAAAATCTGCTTACTCTGAGCGCCTGAAGTTCAATGTGGCGATCAAGATCATCGAC
CGCAAGAAGGCCCCCGCAGACTTCTTGGAGAAATTCCTTCCCCGGGAAATTGAGATTCTGGCCATGTTAA
ACCACTGCTCCATCATTAAGACCTACGAGATCTTTGAGACATCACATGGCAAGGTCTACATCGTCATGGA
GCTCGCGGTCCAGGGCGACCTCCTCGAGTTAATCAAAACCCGGGGAGCCCTGCATGAGGACGAAGCTCGC
AAGAAGTTCCACCAGCTTTCCTTGGCCATCAAGTACTGCCACGACCTGGACGTCGTCCACCGGGACCTCA
AGTGTGACAACCTTCTCCTTGACAAGGACTTCAACATCAAGCTGTCCGACTTCAGCTTCTCCAAGCGCTG
CCTGCGGGATGACAGTGGTCGAATGGCATTAAGCAAGACCTTCTGTGGGTCACCAGCGTATGCGGCCCCA
GAGGTGCTGCAGGGCATTCCCTACCAGCCCAAGGTGTACGACATCTGGAGCCTAGGCGTGATCCTCTACA
TCATGGTCTGCGGCTCCATGCCCTACGACGACTCCAACATCAAGAAGATGCTGCGTATCCAGAAGGAGCA
CCGCGTCAACTTCCCACGCTCCAAGCACCTGACAGGCGAGTGCAAGGACCTCATCTACCACATGCTGCAG
CCCGACGTCAACCGGCGGCTCCACATCGACGAGATCCTCAGCCACTGCTGGATGCAGCCCAAGGCACGGG
GATCTCCCTCTGTGGCCATCAACAAGGAGGGGGAGAGTTCCCGGGGAACTGAACCCTTGTGGACCCCCGA
ACCTGGCTCTGACAAGAAGTCTGCCACCAAGCTGGAGCCTGAGGGAGAGGCACAGCCCCAGGCACAGCCT
GAGACAAAACCCGAGGGGACAGCAATGCAAATGTCCAGGCAGTCGGAGATCCTGGGTTTCCCCAGCAAGC
CGTCGACTATGGAGACAGAGGAAGGGCCCCCCCAACAGCCTCCAGAGACGCGGGCCCAGTGAGCTTCTTG
CGGCC  (SEQ ID NO:7).
```

In a BLASTX search, the full amino acid sequence of the protein of the invention was found to have 307 of 364 amino acid residues (84%) identical to, and 329 of 364 residues (90 %) positive with, the 364 amino acid residue testis-specific serine/threonine kinase from M. musculus (ptnr:SWISSPROT-ACC:Q61241).

The first 70 amino acids of Acc. No. AC010431_A (SER2) were used for signal peptide prediction. The results indicate that the protein likely has no signal peptide, and that it is predicted to localize to the cytoplasm.

A search of the PROSITE database of protein families and domains confirmed that a SER2 polypeptide is a member of the serine/threonine kinase family, which is defined by two signature sequences, as defined above. The SER2 polypeptide sequence contains both Eukaryotic protein kinase signature sequences as defined by the PROSITE database (illustrated by bold and bold, italics in SEQ ID NO:6 of Table 7). The protein kinase ATP binding signature sequence is located at amino acid residues 18-41 of SEQ ID NO:6 (bold). The serine/threonine protein kinase active site signature sequence is located at amino acid residues 132-144 of SEQ ID NO:6 (bold, italic).

Additionally, a search of the Pfam protein families database of alignments shows that SEQ ID NO:6 belongs to the Eukaryotic protein kinase family, identified by Accession Number PF00069.

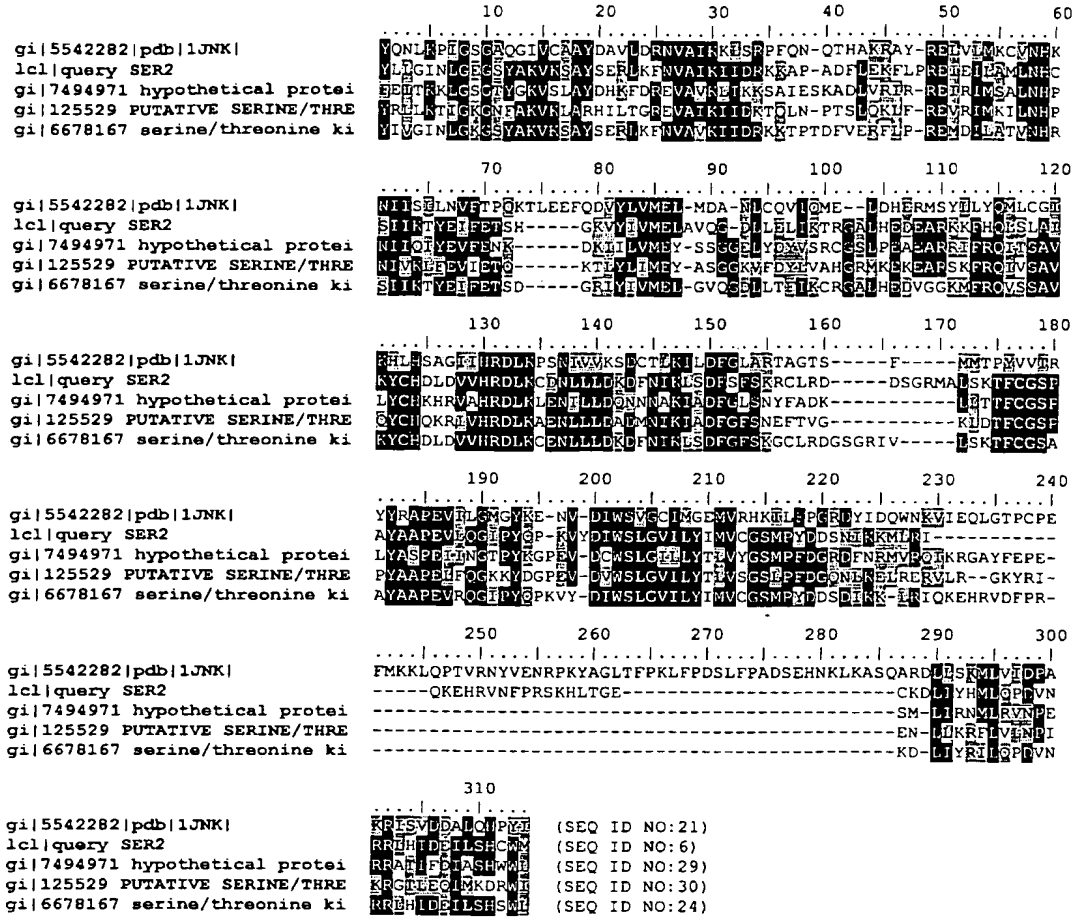
Furthermore, a search of the NCBI CD-Search database demonstrated that SER2 shows significant similarity to three kinase domains, as shown in Table 9.

Table 9: SER2 CDD Domain Homology

Sequences producing significant alignments:		Score (bits)	E value	Row
gnl|Smart|S_TKc	Serine/Threonine protein kinases, catalytic domain...	188	5e-49	1
gnl|Pfam|pfam00069	Pkinase, Eukaryotic protein kinase domain	162	4e-41	2
gnl|Smart|TyrKc	Tyrosine kinase, catalytic domain; Phosphotransferases...	89.7	2e-19	3

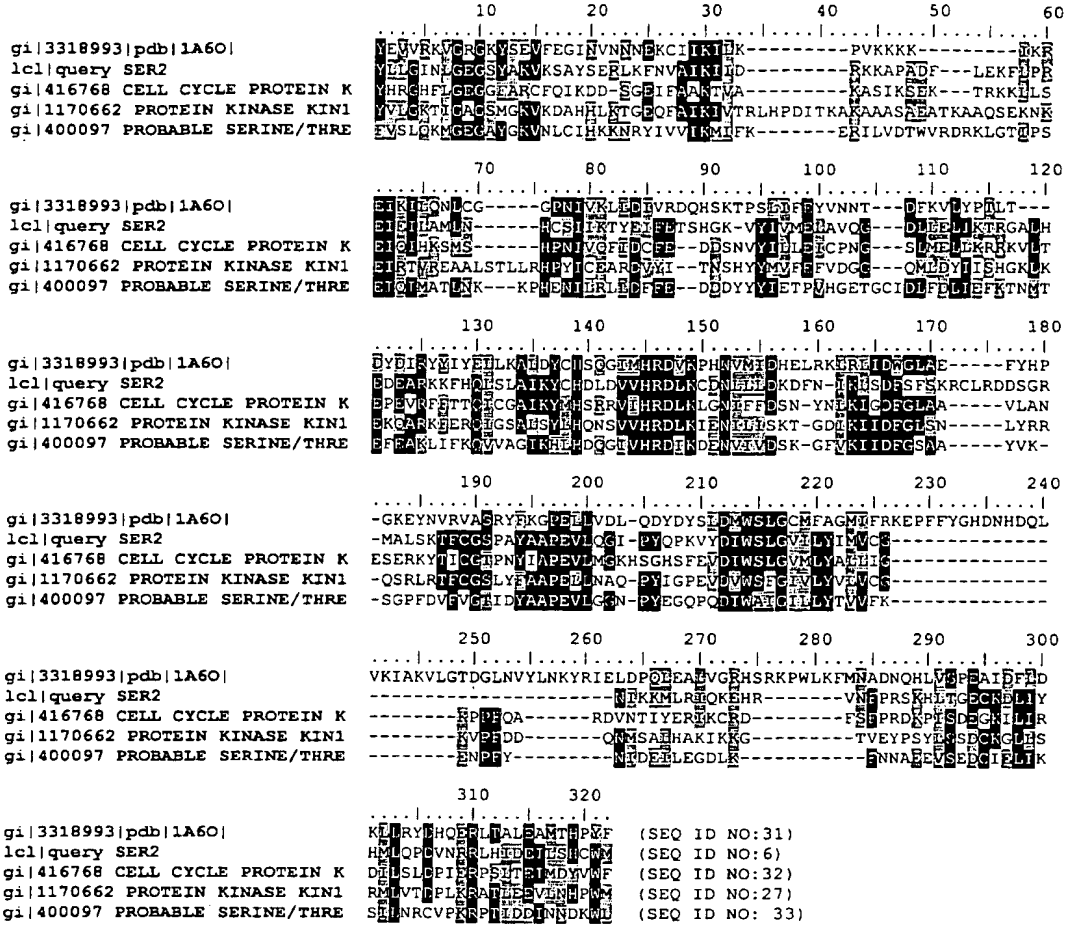
This degree of homology between a SER2 polypeptide and the kinase polypeptide domains (both in terms of length and complexity) is very unlikely to have occurred by chance alone (e.g., the Expect (E) value in Table 8 less than $5e^{-49}$ by chance for the Serine/Threonine protein kinase catalytic domain).

Table 10a: Multiple CD Align- catalytic domain



As illustrated in Table 10A, the SER2 polypeptide has homology to serine/threonine protein kinase catalytic domain (Row 1, Table 9). Entries '1JNK,' '125529,' and '6678167' were discussed above for SER1 multiple alignment (Table 5a). The entry '7494971' corresponds to hypothetical protein B0496.3 (Accession number T29253) from *Caenorhabditis elegans*.

Table 10b: Multiple CD Align- pkinase ATP



The CD multiple Align for SER2 to the pkinase ATP binding domain (Row 2, Table 9) is shown in Table 10b. The figure indicates that SER2 has homology to 'gi 3318993,' and 'gi 1170662,' as discussed above for SER1. Also shown in Table 5b is entry 'gi 416768' (accession number P32562), a yeast serine/threonine protein kinase required for the cell cycle and belongs to the CDC5/MSD2 subfamily. Kitada et al., Mol. Cell. Biol. 13:4445-4457 (1993). Finally, entry 'gi 400097' (accession number P31374) is a probable serine/threonine protein kinase from *S. cerevisiae*. Clark et al., Yeast 9:543-549 (1993).

Based on its relatedness to the serine-threonine kinases, the SER2 protein is a novel member of this protein family. The discovery of molecules related to spermiogenesis and cell cycle regulation satisfies a need in the art by providing new diagnostic or therapeutic

compositions useful in the treatment of disorders associated with alterations in the expression of members of serine/threonine kinase- like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in a variety of diseases and pathologies, including by way of nonlimiting example, those involving hyperproliferative diseases, hypoproliferative diseases, as well as those involving issues relating to fertility.

A SER2 or SER3 nucleic acid is useful for detecting specific cell-types. For example a SER2 or SER3 nucleic acid according to the invention can be present in different levels in various cells. Also, according to the invention the expression of a SER2 or SER3 nucleic acid has utility in identifying tumors. In certain cancers, the appearance or elevation of plasma levels of serine/threonine kinases or increased or decreased phosphorylation of certain proteins are reliable markers of cancer progression. Accordingly, expression levels of serine-threonine kinase- like nucleic acids such as SER2 or SER3 are also useful in diagnosis and prognosis of such cancers.

SER4

A SER4 sequence according to the invention is a nucleotide sequence encoding a polypeptide related to a serine/threonine kinase (STK). A SER4 nucleic acid and its encoded polypeptide includes the sequences shown in Table 11 and 12, respectively. The disclosed nucleic acid sequence (SEQ ID NO:8, Acc NO: AC010761 from GenbankNEW) is 993 nucleotides long. An open reading frame was identified beginning with an initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 991-993. The encoded polypeptide is 330 amino acid residues long (SEQ ID NO:9).

Table 11: Nucleotide sequence including the sequence encoding STK-like protein SER4 (24111358_EXT1) of the invention..

ATGGAAACTGAGGAGCAAGTCCGCGTGAAGGGGAGAGGTGCCTTCGGGATTGTGCAACTGTGCCTGCGAAAGGCTGACCA
GAAGCTGGTGATCATCAAGCAGATTCCAGTGGAACAGATGACCAAGGAAGAGCGGCAGGCAGCCCAGAATGAGTGCCAGG
TCCTCAAGCTGCTCAACCACCCCAATGTCATTGAGTACTACGAGAACTTCCTGGAAGACAAAGCCCTTATGATCGCCATG
GAATATGCACCAGGTGGGACTCTGGCTGAGTTCATCCAAAAGCGCTGTAATTCCCTGCTGGAGGAGGAGACCATCCTGCA
CTTCTTCGTGCAGATCCTGCTTGCACTGCATCATGTGCACACCCACCTCATCCTGCACCGAGACCTCAAGACCCAGAACA
TCCTGCTTGACAAACACCGCATGGTCGTCAAGATCGGTGATTTCGGCATCTCCAAGATCCTTAGCAGCAAGAGCAAGGCC
TACACGGTGGTGGGTACCCCATGCTATATCTCCCCTGAGCTGTGTGAGGGCAAGCCCTACAACCAGAAGAGTGACATCTG
GGCCCTGGGCTGTGTCCTCTACGAGCTGGCCAGCCTCAAGAGGGCTTTCGAGGCTGCGGTGAGTGTATGCACCCTCCAGG
GGACAACTGAGAAATCTACTGCCTCGCCCAGCAGCCCTCTATCTGACCGGTACAGCCCTGAGCTTCGCCAGCTGGTCCTG
AGTCTACTCAGCCTGGAGCCTGCCCAGCGGCCACCACTCAGCCACATCATGGCACAGCCCCTCTGCATCCGTGCCCTCCT
CAACCTCCACACCGACGTGGGCAGTGTCCGCATGCGGAGGCCTGTGCAGGGACAGCGAGCGGTCCTGGGCGGCAGGGTGT
GGGCACCCAGTGGGAGCACACTTTCGCCTCTGACTGTGTCCGCCACAGCCTGCACCTACACTCTGTCATCTTTTACCATT
GACACCTTGCACCATGATCTGAAAACACAA**TGA** (SEQ ID NO:8).

Table 12: Protein sequence encoded by SER4.

METEEQVRVKGRGAFGIVQLCLRKADQKLVIIKQIPVEQMTKEERQAAQNECQVLKLLNHPNVIEY
YENFLEDKALMIAMEYAPGGTLAEFIQKRCNSLLEEETILHFFVQILLALHHVHTHL**ILHRDLKTQN**
ILLDKHRMVVKIGDFGISKILSSKSKAYTVVGTPCYISPELCEGKPYNQKSDIWALGCVLYELASLK
RAFEAAVSVCTLQGTTEKSTASPSSPLSDRYSPELRQLVLSLLSLEPAQRPPLSHIMAQPLCIRALLNL
HTDVGSVRMRRPVQGQRAVLGGRVWAPSGSTLSPLTVSATACTYTLSSFTIDTLHHDLKTQ
(SEQ ID NO:9).

Based on information available from expression data, as well as the expression of STK family members, it is likely that 24111358_EXT1 (SER4) is expressed in fetal lung, other developmental tissues and sex tissues.

The hydropathy profile for SER4 indicates that this sequence has a strong signal peptide toward the 5' terminal supporting extracellular localization. The PSORT and SignalP profile also offer significant supportive evidence. It is very likely a membrane-bound peptide as the protein predicted here is similar to the STK gene family, some members of which are localized at the plasma membrane. Therefore it is likely that this novel gene is available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of SER4 has 429 of 699 nucleotides (61%) identical to Homo sapiens protein kinase SID6-1512 (Genbank ID: AB026289). In a BLASTX identity search the full amino acid sequence of the protein of the invention was found to have 150 of 333 amino acid residues (45%) identical to, and 213 of 333 amino acid residues (63%) positive with, the 357 amino

acid residue protein similar to the CDC2/CDX subfamily of ser/thr protein kinases from Caenorhabditis elegans (SPTREMBL-ACC:O01775).

A search of the Prosite database revealed the occurrence of the serine/threonine protein kinase active site signature sequence at residues 124-136 (shown in bold) of the SER4 polypeptide (SEQ ID NO:9).

Alignment in the Pfam database showed similarity of amino acids 4-262 to the pkinase domain, with an E value of $8.4e^{-71}$. This sequence therefore belongs to the Eukaryotic protein kinase family, identified by Accession Number PF00069.

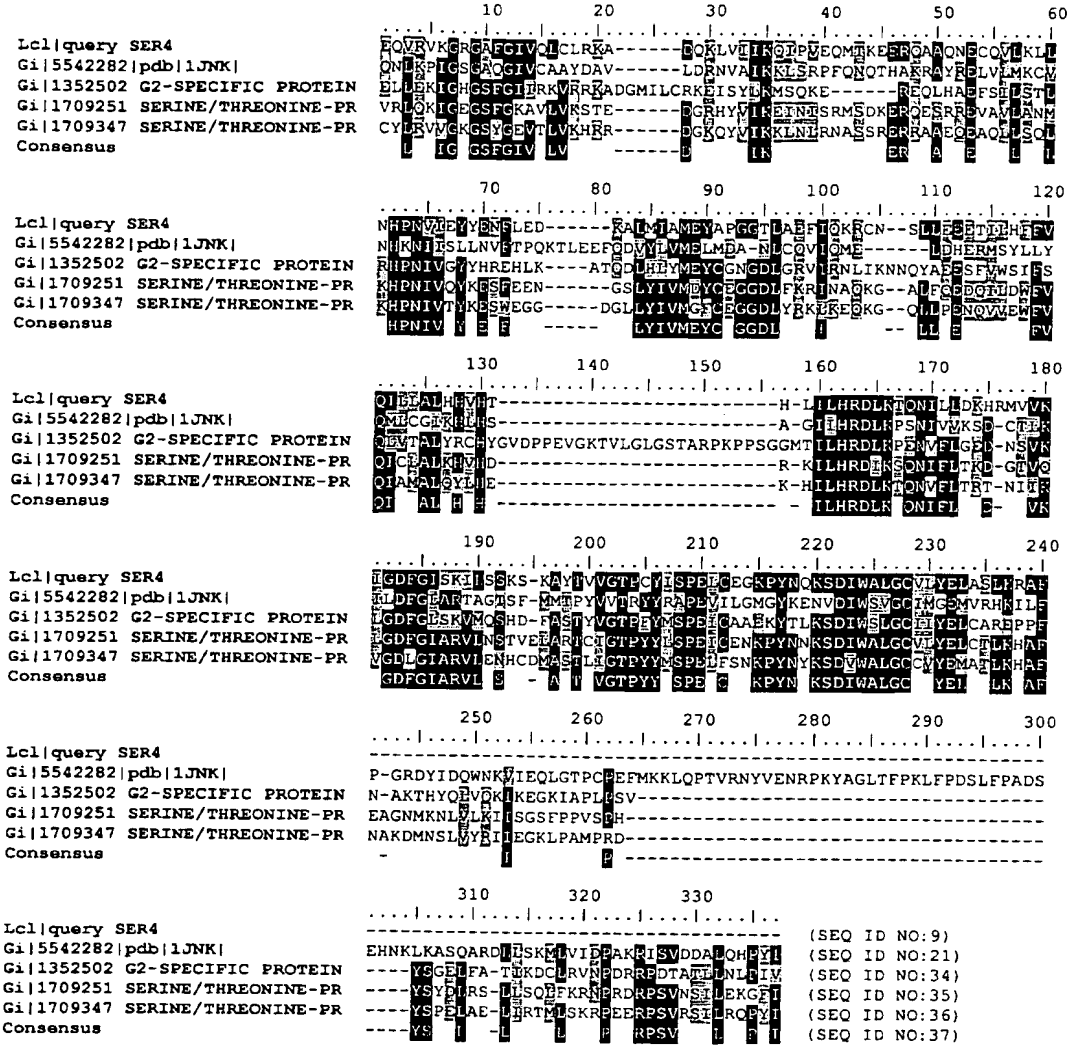
Furthermore, a search of the NCBI CD-Search database demonstrated that SER4 shows significant similarity to three kinase domains, as shown in Table 13.

Table 13: SER4 CDD Domain Homology

Sequences producing significant alignments:		**Score (bits)**	**E value**
gnl|Smart|S_TKc	Serine/Threonine protein kinases, catalytic domain...	197	9e-52
gnl|Pfam|pfam00069	Pkinase, Eukaryotic protein kinase domain	185	2e-48
Gnl|Smart|TyrKc	Tyrosine kinase, catalytic domain; Phosphotransferases...	116	1e-27

This degree of homology between a SER4 polypeptide and the kinase polypeptide domains (both in terms of length and complexity) is very unlikely to have occurred by chance alone (e.g., the Expect (E) value in Table 13 less than $9e^{-52}$ by chance for the serine/threonine protein kinase catalytic domain).

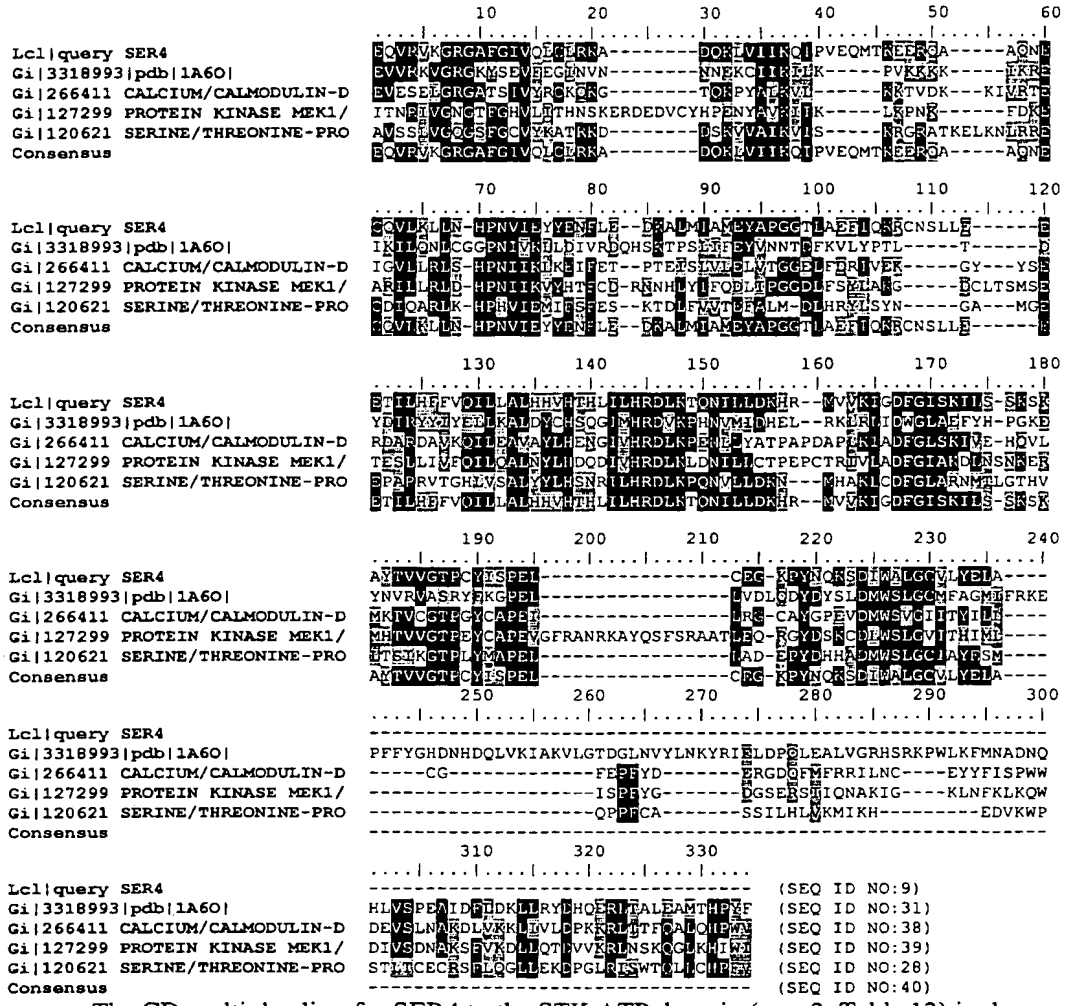
Table 14a: Multiple CD Align- catalytic domain



The CD multiple align for SER4 to the STK catalytic domain (row 1, Table13) is shown in Table 14a. The figure indicates that SER4 has homology to 1JNK (accession number 5542282), as discussed above. Also shown in Table 14a is entry '1352502 (accession number P48479), a G2-specific serine/threonine kinase, NIM-1 from *Neurospora crassa*. Pu et al., J. Biol. Chem. 270:18110-18116 (1995). Entry '1709251' (accession number P51954) is serine/threonine protein kinase NEK1 (NEMA-related protein kinase-1) from mouse, which is highly expressed in meiotic germ cells. Mizzen et al., EMBO J. 11:3521-3531 (1992). Finally, entry '1709347' (accession number P51957) is serine/threonine kinase NRK2 from

human breast tissue, which is related to cell cycle regulating kinases. Levedakou et al., Oncogene 9:1977-1988 (1994).

Table 14b: Multiple CD Align- pkinase ATP domain



The CD multiple align for SER4 to the STK ATP domain (row 2, Table 13) is shown in Table 14b. The figure indicates that SER4 has homology to 'gi 3318993' and 'gi120621,' as discussed above. Also shown in Table 14b is entry '266411' (accession number P08414), a mouse calcium-calmodulin dependent (serine/threonine) protein kinase, found in brain and testes. Jones et al., FEBS Lett. 289:105-109 (1991). Also, entry '127299' (accession number P24719) is serine/threonine kinase MEK1/MRE4 from yeast, which is required for meiotic recombination. Leem & Ogawa, Nucl. Acids Res. 20:449-457 (1992).

The expression pattern, map location and protein similarity information for the sequence of Acc. No. 24111358_EXT1 suggest that this gene functions as a member of the "STK-like protein family". Specifically, SER4 has similarity to members of the CDC2/CDX subfamily of the STK proteins and may be involved in regulating mitosis. For example, CDC2 is a catalytic subunit of a protein kinase complex called the M-phase promoting factor, which induces entry into mitosis, and is universal among eukaryotes.

The SER4 may be potentially used as a protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, for gene therapy (gene delivery/gene ablation), as research tools, for tissue regeneration *in vitro* and *in vivo* (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues).

The nucleic acids and proteins provided by SER4 are useful in potential therapeutic applications implicated, for example but not limited to, the following diseases and disorders: Systemic lupus erythematosus, Autoimmune disease, Asthma, Emphysema, Scleroderma, Cancer, Fertility disorders, Reproductive disorders, Tissue/Cell growth regulation disorders, Developmental disorders and resulting disorders derived from the above conditions, and/or other pathologies/disorders.

For example, a cDNA encoding STK-like protein may be useful in gene therapy, and the STK-like protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the pathologies described above. The novel nucleic acid encoding the STK-like protein, and the STK-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

SER4, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The SER4 polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

SER5

A SER5 sequence according to the invention is a nucleotide sequence encoding a polypeptide related to a serine protease hepsin-like protein.

A SER5 nucleic acid and its encoded polypeptide includes the sequences shown in Table 15 and 16, respectively. The disclosed nucleic acid sequence (SEQ ID NO:10, Acc NO: GM_10221687) is 1264 nucleotides long. An open reading frame was identified beginning with an initiation ATG codon at nucleotides 8-10 and ending with a TAA codon at nucleotides 1256-1258. The encoded polypeptide is 416 amino acid residues long (SEQ ID NO:11).

Table 15: Nucleotide sequence including the sequence encoding serine protease-like protein SER5 (GM_10221687) of the invention..

AAGGCCA**ATG**ACACTGGGTAGAAGAGTGAGTTCACTGAAACCATGGATGTTTGCCCTTATTGTCAGAGCT
GTTGTGTTGATTCTGGTGATACTGATTGGTCTCCTTGTTTATTTTTTGGCATATGGCCTGAAGTTTTACT
ATTACCAGACCTCCTTCCAGATCCCCAGTATTGAATATAATCCTGATTTTTCAGTAGAACACTCAAAACT
TAGCACCGACCTGAAACAAAAAGTCAGTAACGAGATATTTCAGAGATCCAATTTAAACCATCATTACATA
AAGTGTCAAGTTGTCAACTTTAGAGTCCCAGAGGAAGATGGTGTGAAAGTAGATGTCATTATGGTGTTCC
AGTTCCCCTCTACTGAACAAAGGGCAGTAAGAGAGAAGAAAATCCAAAGCATCTTAAATCAGAAGATAAG
GAATTTAAGAGCCTTGCCAATAAATGCCTCATCAGTTCAAGTTAATGGTAAGTTAACTGTCCAAGCAATC
TCATCTTTTTCAGGTTGTGGTAAACGAGTTGTTCCATTAAACGTCAACAGAATAGCATCTGGAGTCATTG
CACCCAAGGCGGCCTGGCCTTGGCAAGCTTCCCTTCAGTATGATAACATCCATCAGTGTGGGGCCACCTT
GATTAGTAACACATGGCTTGTCACTGCAGCACACTGCTTCCAGAATTTTTCCAGGTATAAAAATCCACAT
CAATGGACTGTTAGTTTTGGAACAAAAATCAACCCTCCCTTAATGAAAAGAAATGTCAGAAGATTTATTA
TCCATGAGAAGTACCGCTCTGCAGCAAGAGAGTACGACATTGCTGTTGTGCAGGTCTCTTCCAGAGTCAC
CTTTTCGGATGACATACGCCGGATTTGTTTGCCAGAAGCCTCTGCATCCTTCCAACCAAATTTGACTGTC
CACATCACAGGATTTGGAGCACTTTACTATGGTGGGGAATCCCAAAATGATCTCCGAGAAGCCAGAGTGA
AAATCATAAGTGATGATGTCTGCAAGCAACCACAGGTGTATGGCAATGATATAAAACCTGGAATGTTCTG
TGCCGGATATATGGAAGGAATTTATGATGCCTGCAGGGGTGATTCTGGGGGACCTTTAGTCACAAGGGAT
CTGAAAGATACGTGGTATCTCATTGGAATTGTAAGCTGGGGAGATAACTGTGGTCAAAAGGACAAGCCTG
GAGTCTACACACAAGTGACTTATTACCGAAACTGGATTGCTTCAAAAACAGGCATC**TAA**AATAAA (SEQ ID NO:10)

Table 16. Protein sequence encoded by SER5.

MTLGRRVSSLKPWMFALIVRAVVLILVILIGLLVYFLAYGLKFYYYQTSFQIPSIEYNPDFSVEHSKLST
DLKQKVSNEIFQRSNLNHHYIKCQVVNFRVPEEDGVKVDVIMVFQFPSTEQRAVREKKIQSILNQKIRNL
RALPINASSVQVNGKLTVQAISSFSGCGKRVVPLNVNRIASGVIAPKAAWPWQASLQYDNIHQCGATLIS
NTWL**VTAAHC**FQNFSRYKNPHQWTVSFGTKINPPLMKRNVRRFIIHEKYRSAAREYDIAVVQVSSRVTFS
DDIRRICLPEASASFQPNLTVHITGFGALYYGGESQNDLREARVKIISDDVCKQPQVYGNDIKPGMFCAG
YMEGIY***DACRGDSGGPLV***TRDLKDTWYLIGIVSWGDNCGQKDKPGVYTQVTYYRNWIASKTGI (SEQ ID NO:11).

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of SER5 has 749 of 1223 bases (61%) identical to a *Homo sapiens* serine protease hepsin mRNA (Patent publication AU9527248-A published 08-FEB-1996). The full amino

acid sequence of the protein of the invention was found to have 181 of 420 amino acid residues (43%) identical to, and 266 of 420 residues (63%) positive with, the 422 amino acid residue Serine Protease DESC1 protein from *Homo sapiens* (ptnr: TREMBLNEW-ACC:AAF04328).

Hepsin is a putative membrane-bound serine protease, was originally identified as a human liver cDNA clone. The messenger RNA of hepsin is 1.85 kilobases in size and present in most tissues, with the highest level in liver. Hepsin is synthesized as a single polypeptide chain, and its mature form of 51 kDa was found in various mammalian cells including HepG2 cells and baby hamster kidney cells. It is present in the plasma-membrane in a molecular orientation of type II membrane-associated proteins, with its catalytic subunit (carboxyl-terminal half) at the cell surface, and its amino terminus facing the cytosol. Hepsin is found neither in cytosol nor in culture media. It is postulated that hepsin has an important role in cell growth and function.

Hepsin is a type II transmembrane serine protease that is highly expressed on the surface of hepatocytes. The physiological function of hepsin is not known, although in vitro studies indicate that hepsin plays a role in the initiation of blood coagulation and in hepatocyte growth, but is not essential for embryonic development and normal hemostasis.

Extracellular proteases mediate the digestion of neighboring extracellular matrix components in initial tumor growth, allow shedding or desquamation of tumor cells into the surrounding environment, provide the basis for invasion of basement membranes in target metastatic organs, and are required for release and activation of many growth and angiogenic factors. Overexpression of the serine protease hepsin gene has been identified in ovarian carcinomas. Torres-Rosado, et al., Proc Natl Acad Sci U S A. ;90:7181-7185 (1993). Quantitative PCR was used to determine the relative expression of hepsin compared to that of beta-tubulin. The mRNA expression levels of hepsin were significantly elevated in 7 of 12 low malignant potential tumors and in 27 of 32 carcinomas. On Northern blot analysis, the hepsin transcript was abundant in carcinoma but was almost never expressed in normal adult tissue, including normal ovary. These results suggest that hepsin is frequently overexpressed in ovarian tumors and therefore may be a candidate protease in the invasive process and growth capacity of ovarian tumor cells.

PSORT analysis predicts the protein encoded by SER5 to be localized extracellularly with a certainty of 0.8200. Using the SIGNALP analysis, it is predicted that the protein of the invention has a signal peptide with most likely cleavage site between pos. 40 and 41: AYG-LK.

A search of the PROSITE database of protein families and domains confirmed that a SER5 polypeptide is a member of the trypsin family of serine proteases. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen- bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases. Brenner, Nature 334:528-530 (1988); Sprang et al., Science 237:905-909 (1987). One conserved region is the histidine active site and the other is the serine active site. Proteins that have both signature sequences have an extremely high probability of being a trypsin family serine protease.

The histidine active site consensus pattern is: [LIVM]-[ST]-A-[STAG]-H-C (SEQ ID NO:12). This pattern is found in amino acids 215-220 of SEQ ID NO:11 (shown in bold).

The serine active site consensus pattern is: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]- [LIVMFYSTANQH] (SEQ ID NO:13). This pattern is indicated by the bold, italic residues 357-368 of SEQ ID NO:11.

The majority of known trypsin family serine proteases belong to the class detected by these patterns. The SER5 polypeptide sequence contains both trypsin family sequences as defined by the PROSITE database (illustrated by bold and bold, italics in SEQ ID NO:11 of Table 16).

A partial list of proteases known to belong to the trypsin family includes: Acrosin; Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C; Cathepsin G; Chymotrypsins; Complement components C1r, C1s, C2, and complement factors B, D and I; Complement-activating component of RA-reactive factor; Cytotoxic cell proteases (granzymes A to H); Duodenase I; Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin); Enterokinase (EC 3.4.21.9) (enteropeptidase); Hepatocyte growth factor activator; Hepsin; Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin); Plasma kallikrein; Mast cell proteases (MCP) 1 (chymase) to 8; Myeloblastin (proteinase 3) (Wegener's autoantigen);

Plasminogen activators (urokinase-type, and tissue-type); Trypsins I, II, III, and IV; Tryptases; Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator; Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab; Apolipoprotein(a); Blood fluke cercarial protease; Drosophila trypsin like proteases: alpha, easter, snake-locus; Drosophila protease stubble (gene sb); Major mite fecal allergen Der p III.

All the above proteins belong to family S1 in the classification of peptidases and originate from eukaryotic species. Rawlings and Barrett Meth. Enzymol. 244:19-61(1994). It should be noted that bacterial proteases that belong to family S2A are similar enough in the regions of the active site residues that they can be picked up by the same patterns. These proteases include Achromobacter lyticus protease I; Lysobacter alpha-lytic protease; Streptogrisin A and B (Streptomyces proteases A and B); Streptomyces griseus glutamyl endopeptidase II; Streptomyces fradiae proteases 1 and 2.

Additionally, a search of the Pfam protein families database of alignments shows that SEQ ID NO:11 belongs to the trypsin family of proteolytic enzymes, identified by Accession Number PF00089. The E value for the alignment of residues 179-407 of SER5 with the trypsin domain is $7e^{-86}$.

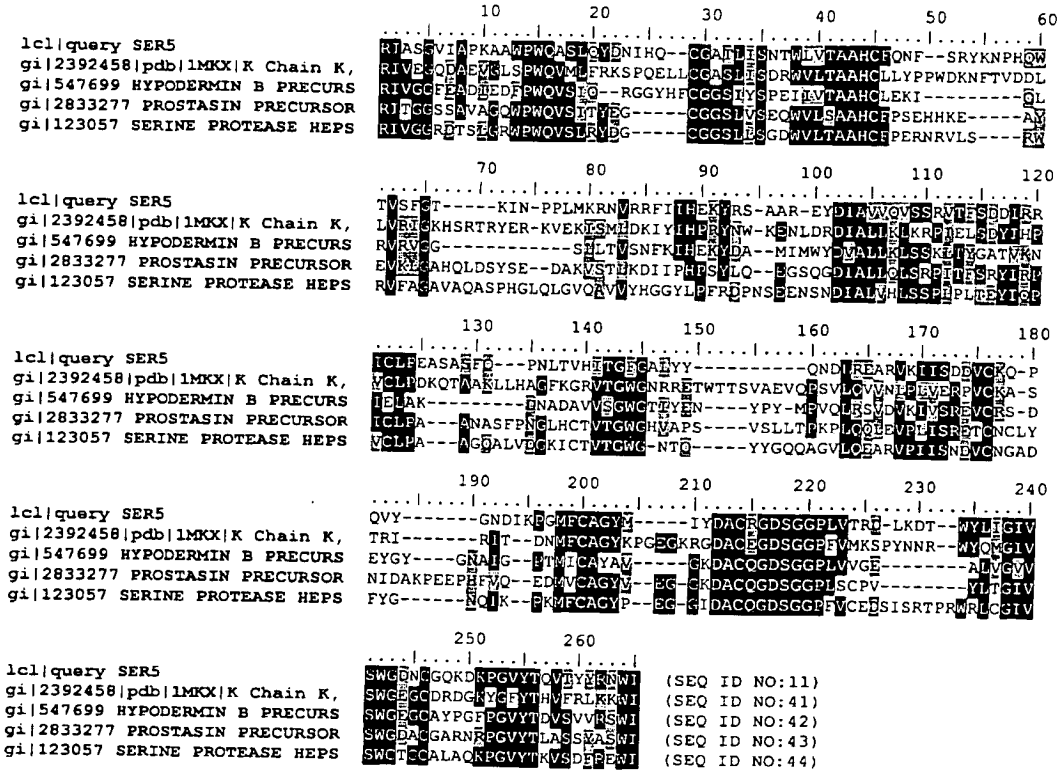
Furthermore, a search of the NCBI CD-Search database demonstrated that SER5 shows significant similarity to two trypsin-like domains, as shown in Table 17.

Table 17: SER5 CDD Domain Homology

Sequences producing significant alignments:		Score (bits)	E value	Row
Gnl|Smart|Tryp_SPc	Trypsin-like serine protease	187	1e-48	1
gnl|Pfam|pfam00089	Trypsin, Trypsin	154	9e-39	2

This degree of homology between a SER5 polypeptide and the kinase polypeptide domains (both in terms of length and complexity) is very unlikely to have occurred by chance alone (e.g., the Expect (E) value in Table 17 less than $1e^{-48}$ by chance for the trypsin-like serine proteases).

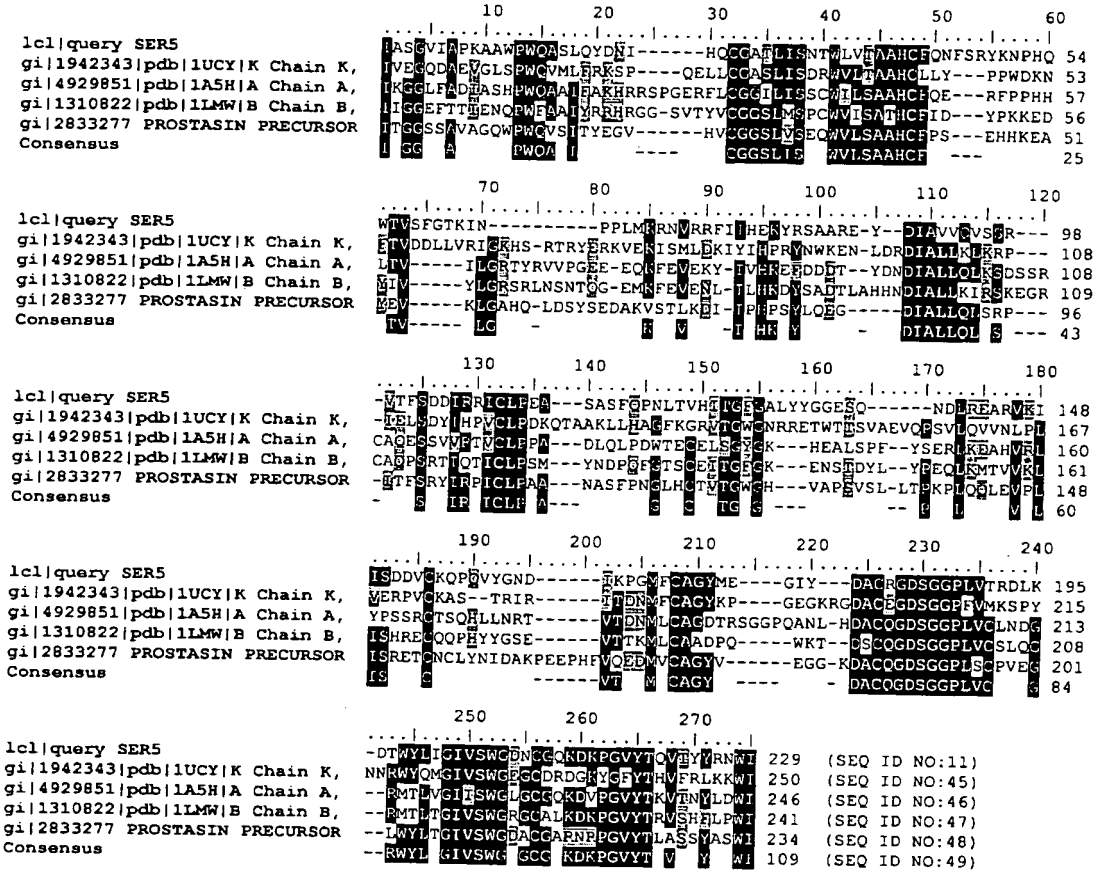
Table 18a: Multiple CD Align-Trypsin-like serine protease



The CD multiple Align for SER5 to trypsin-like serine protease domains (row 1, Table 17) is shown in Table 18a. Many trypsin-like serine proteases are synthesized as inactive precursor zymogens that are cleaved during limited proteolysis to generate their active forms. A few, however, are active as single chain molecules, and others are inactive due to substitutions of the catalytic triad residues. The figure indicates that SER5 has homology to '1MKX_K' (accession number 2392458), a bovine alpha thrombin. Vijayalakshmi et al., Protein Sci. 3:2254-2271 (1994); Martin et al., Biochemistry 35:13030-13039 (1996). Also shown in Table 18a is entry '547699' (accession number P35588), a Hypodermin B Precursor (HB). Lecroisey et al., Eur. J. Biochem. 134:261-267 (1983). Hypodermin B is a trypsin related enzyme found in the insect *Hypoderma lineatum*. Entry '2833277' (accession number Q16651), a human prostatin precursor, which is a human serine proteinase purified from seminal fluid, and is a member of the trypsin family. Yu et al., J Biol. Chem. 270:13483-

13489 (1995). Finally, entry '123057' (accession number P05981) is a human serine protease hepsin (transmembrane protease, serine-1), a type-II membrane protein belonging to the trypsin family and is related to cell growth and maintenance of cell morphology. Leytus et al., Biochemistry 27:1067-1074 (1988); Tsuji et al., J. Biol. Chem. 266:16948-16953 (1991); Torres-Rosado et al., Proc. Natl. Acad. Sci., USA 90:7181-7185 (1993).

Table 18b: Multiple CD Align-Trypsin-family domain



The CD multiple Align for SER5 to trypsin-like domains (peptidase family S1, row 2, Table 17) is shown in Table 18b. The figure indicates that SER5 has homology to '1UCY_K' (accession number 1942343), bovine chain K from thrombin. Martin et al., Biochem. 35:13030-13039 (1996). Also shown in Table 18b is entry 'gi 4929851' (accession number 4929851), a Chain A catalytic domain of human tissue plasminogen activator. Bode and Renatus, Curr. Op. Struct. Biol. 7:21713-21719 (1997). Entry 'gi 1310822' (accession number

1310822) is a human chain B urokinase-type plasminogen activator. Spraggon et al., structure 3:681-691 (1995). Finally, entry ' gi 2833277' (accession number g2833277) is a human prostatin precursor, a membrane-bound protein with trypsin-like cleavage specificity, and is found in the prostate, liver, salivary gland, kidney, lung, pancreas, colon, bronchus, and renal proximal tubular cells. Yu, et al., J. Biol. Chem. 270:13483-13489.

The similarity of SER5 to the proteins of the serine protease family indicates that it is a novel member of this class of proteins. Therefore, the novel nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in various diseases involving blood coagulation, human liver cells, hepatoma cells and hepatocellular carcinoma and/or other pathologies and disorders. For example, a cDNA encoding the Serine Protease (SP) -like protein (e.g., hepsin-like protein, or trypsin like protein) may be useful in gene therapy, and the SP -like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from various diseases involving blood coagulation, human liver cells, hepatoma cells and hepatocellular carcinoma. The novel nucleic acid encoding Serine Protease -like protein, and the SP-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

SER6

A SER6 sequence according to the invention is a nucleotidc sequence encoding a polypeptide related to a serine protease kallikrein-like protein. A SER6 nucleic acid and its

encoded polypeptide includes the sequences shown in Table 19 and 20, respectively. The disclosed nucleic acid sequence (SEQ ID NO:14, Acc NO: 12996895_1) is 1314 nucleotides long. An open reading frame was identified beginning with an initiation ATG codon at nucleotides 1-3 and ending with a stop codon at nucleotides 1264-1266. The encoded polypeptide is 421 amino acid residues long (SEQ ID NO:15).

Table 19: Nucleotide sequence including the sequence encoding serine protease-like protein SER6 (12996895_1) of the invention..

```
ATGGAGAGCCCAGGTACGAGCCTGCCCAAGTTCACCTGGCGGGAGGGCCAGAAGCAGCTACCGCTCATCGGGTGCGTGCT
CCTCCTCATTGCCCTGGTGGTTTCGCTCATCATCCTCTTCCAGTTCTGGCAGGGCCACACAGGGATCAGGTACAAGGAGC
AGAGGGAGAGCTGTCCCAAGCACGCTGTTCGCTGTGACGGGGTGGTGGACTGCAAGCTGAAGAGTGACGAGCTGGGCTGC
GTGAGGTTTGACTGGGACAAGTCTCTGCTTAAAATCTACTCTGGGTCCTCCCATCAGTGGCTTCCCATCTGTAGCAGCAA
CTGGAATGACTCCTACTCAGAGAAGACCTGCCAGCAGCTGGGTTTCGAGAGTGCTCACCGGACAACCGAGGTTGCCCACA
GGGATTTTGCCAACAGCTTCTCAATCTTGAGATACAACTCCACCATCCAGGAAAGCCTCCACAGGTCTGAATGCCCTTCC
CAGCGGTATATCTCCCTCCAGTGTTCCCACTGCGGACTGAGGGCCATGACCGGGCGGATCGTGGGAGGGGCGCTGGCCTC
GGATAGCAAGTGGCCTTGGCAAGTGAGTCTGCACTTCGGCACCACCCACATCTGTGGAGGCACGCTCATTGACGCCCAGT
GGGTGCTCACTGCCGCCCACTGCTTCTTCGTGACCCGGGAGAAGGTCCTGGAGGGCTGGAAGGTGTACGCGGGCACCAGC
AACCTGCACCAGTTGCCTGAGGCAGCCTCCATTGCCGAGATCATCATCAACAGCAATTACACCGATGAGGAGGACGACTA
TGACATCGCCCTCATGCGGCTGTCCAAGCCCCTGACCCTGTCCGCTCACATCCACCCTGCTTGCCTCCCCATGCATGGAC
AGACCTTTAGCCTCAATGAGACCTGCTGGATCACAGGCTTTGGCAAGACCAGGGAGACAGATGACAAGACATCCCCCTTC
CTCCGGGAGGTGCAGGTCAATCTCATCGACTTCAAGAAATGCAATGACTACTTGGTCTATGACAGTTACCTTACCCCAAG
GATGATGTGTGCTGGGGACCTTCGTGGGGGCAGAGACTCCTGCCAGGGAGACAGCGGGGGGCCTCTTGTCTGTGAGCAGA
ACAACCGCTGGTACCTGGCAGGTGTCACCAGCTGGGGCACAGGCTGTGGCCAGAGAAACAAACCTGGTGTGTACACCAAA
GTGACAGAAGTTCTTCCCTGGATTTACAGCAAGATGGAGAGCGAGGTGCGATTCACAAAATCCTAACCAGCTGGCCTGCT
GCTCTGCACAGCACCGGCTGCTGTGAAGACTCTG  (SEQ ID NO:14).
```

Table 20 Protein sequence encoded by SER6.

```
MESPGTSLPKFTWREGQKQLPLIGCVLLLIALVVSLIILFQFWQGHTGIRYKEQRESCPKHAVRCDGVVDCKLKSDELGCV
RFDWDKSLLKIYSGSSHQWLPICSSNWNDSYSEKTCQQLGFESAHRTTEVAHRDFANSFSILRYNSTIQESLHRSECPSQR
YISLQCSHCGLRAMTGRIVGGALASDSKWPWQVSLHFGTTHICGGTLIDAQWVLTAAHCFFVTREKVLEGWKVYAGTSNLH
QLPEAASIAEIIINSNYTDEEDDYDIALMRLSKPLTLSAHIHPACLPMHGQTFSLNETCWITGFGKTRETDDKTSPFLREV
QVNLIDFKKCNDYLVYDSYLTPRMMCAGDLRGGRDSCQGDSGGPLVCEQNNRWYLAGVTSWGTGCGQRNKPGVYTKVTEVL
PWIYSKMESEVRFTKS  (SEQ ID NO:15)
```

In a search of sequence databases, it was found, for example, that the protein encoded by CuraGen Acc. No. 12996895_1 (SER6) has 150 of 396 amino acid (37 %) identity to *Mus musculus* mosaic serine protease epithelisasin protein (ACC: AAF21308). The full amino acid sequence of the protein of the invention was also found to have 153 of 400 amino acid residues (38%) identical to, and 225 of 400 residues (56%) positive with, the 492 amino acid residue transmembrane protease, serine 2 (EC 3.4.21.) protein from Homo sapiens (ptnr:SPTREMBL-ACC:O15393). Thus, the protein of Acc. No. 12996895_1 is a homolog of proteins in the Kallikrein family.

Recker et al., Urology 55:481-485 (2000), describes human glandular kallikrein as a tool to improve discrimination of poorly differentiated and non-organ-confined prostate cancer

compared with prostate-specific antigen. Human glandular kallikrein possesses 80% structure identity with prostate-specific antigen and is secreted by identical prostate epithelial cells. Although increasing with pathologic stage, prostate-specific antigen is not clinically sufficient to predict histologic grade and pathologic stage of prostate cancer in individual cases. Glandular kallikrein was helpful in the prediction of organ-confined disease, and may serve a useful tool for more accurate prediction of tumor grade or stage and allow better clinical decision-making.

PSORT analysis predicts the protein of CuraGen Acc. No. 12996895_1 to be localized outside the cell with a certainty of 0.82. Using the SIGNALP analysis, it is predicted that the protein of the invention has a cleavable N-terminal signal sequence with a most likely cleavage site between pos. 48 and 49: HTG-IR. The predicted molecular weight of the protein of the invention is 474840.5 daltons. The protein is expressed in osteosacroma cells.

A search of the PROSITE database of protein families and domains confirmed that a SER6 polypeptide is a member of the trypsin family of serine proteases. The histidine active site consensus pattern is: [LIVM]-[ST]-A-[STAG]-H-C (SEQ ID NO:12). This pattern is found in amino acids 216-221 of SEQ ID NO:15 (shown in bold, Table 20).

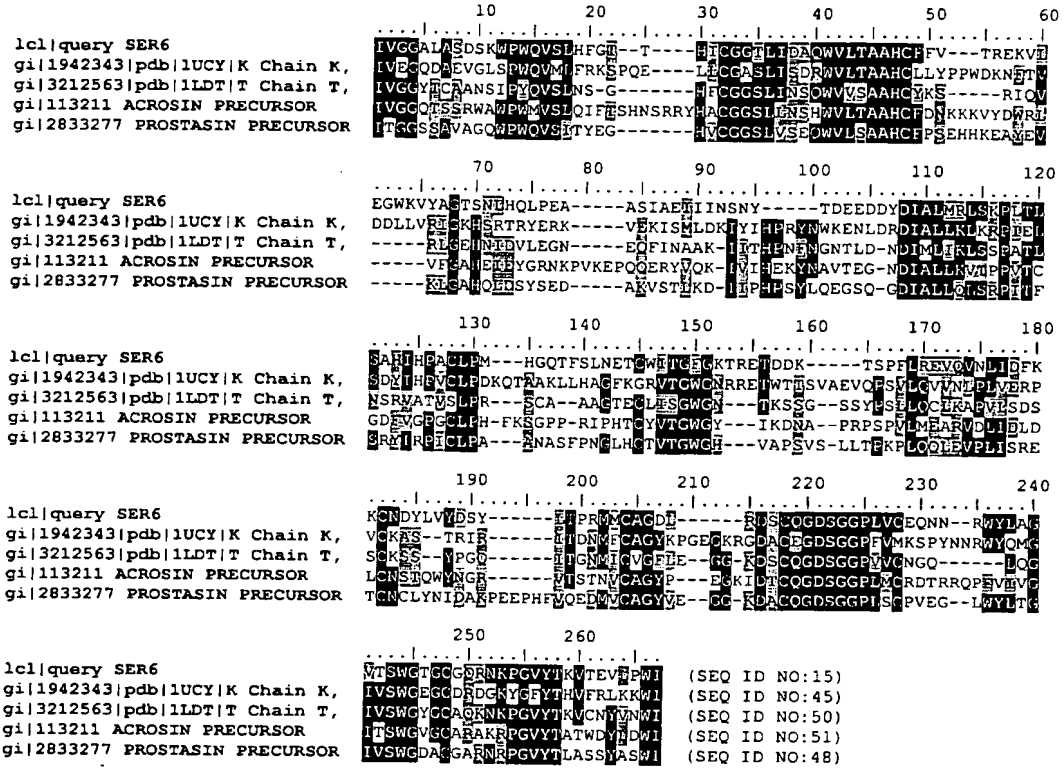
The serine active site consensus pattern is: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]- [LIVMFYSTANQH] (SEQ ID NO:13). This pattern is indicated by the bold, italic residues 359-370 of SEQ ID NO:15 in Table20.

Additionally, a search of the Pfam protein families database of alignments shows that SEQ ID NO:14 belongs to the trypsin family of proteolytic enzymes, identified by Accession Number PF00089. The E value for the alignment of residues 180-408 of SER6 with the trypsin domain is $1.9e^{-90}$.

Table 21: SER6 CDD Domain Homology

Sequences producing significant alignments:		Score (bits)	E value	Row
gnl|Pfam|pfam00089	Trypsin, Trypsin	168	4e-43	1
gnl|Smart|Tryp_SPc	Trypsin-like serine protease	164	6e-42	2

Table 22a: Multiple CD Align trypsin-like domain



The CD multiple Align for SER6 is shown in Table 22a. The figure indicates that SER6 has homology to '1UCY_K' bovine chain k thrombin (see above) and prostatin precursor, entry '2833277,' a human prostatin precursor, and a member of the trypsin family, which is a human serine proteinase purified from seminal fluid. Yu et al., J Biol. Chem. 270:13483-13489 (1995). Also shown in Table 22a is entry '1LDT_T' (accession number 3212563), a leech-derived human tryptase inhibitor. Sommerhoff, et al., Biol. Chem. Hoppe Seyler 375:685-694 (1994); Auerswald, et al. Biol. Chem. Hoppe Seyler 375:695-703(1994);

Muhlhahn, et al., FEBS Lett. 355:290-296 (1994); and Stubbs, et al., J. Biol. Chem. 272:19931-19937 (1997). Entry '113211' (accession number P29293) is a precursor of Acrosin, the major protease of mammalian spermatozoa, and is a serine protease with trypsin-like cleavage specificity. Klemm et al., Biochim. Biophys Acta. 1090:270-272 (1991).

Table 22b: Multiple CD Align serine-protease (trypsin-like) domain

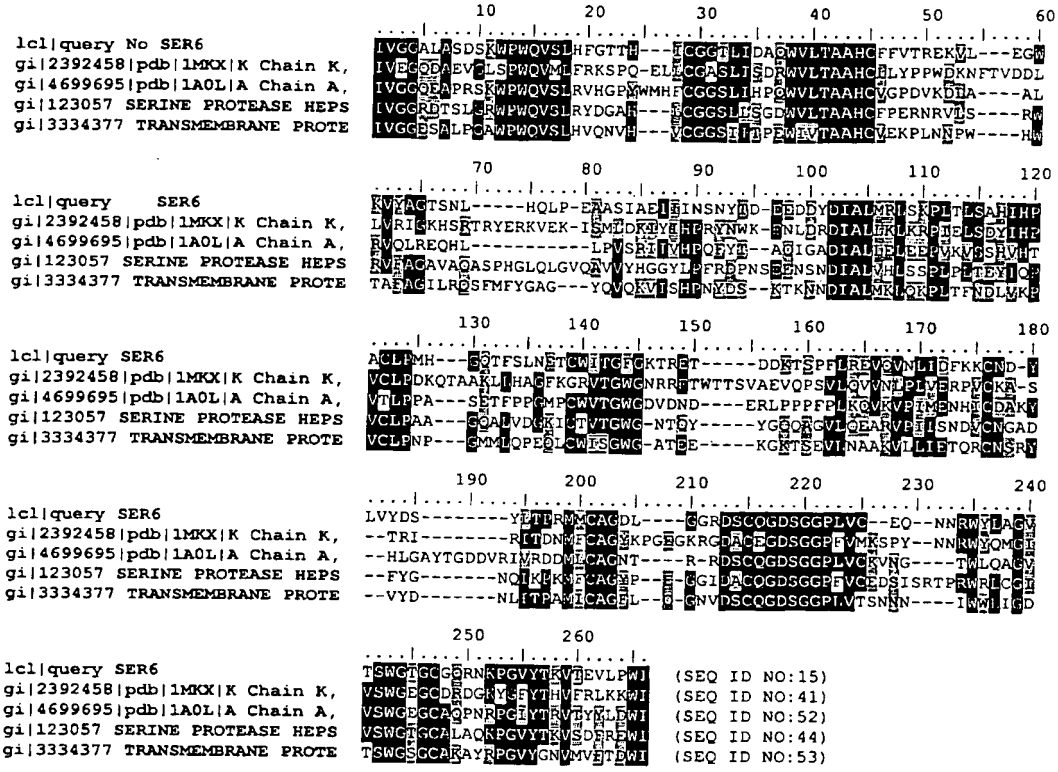


Figure 22b indicates that SER6 has homology to '1MKX_K' (accession number 2392458), as discussed above. Entry '123057' a human serine protease hepsin was also discussed above. Also shown in Table 22b is entry '4699695' (accession number 4699695), a chain A human beta tryptase. Periera et al., Nature 392:306-311 (1998). Finally, entry 'gi 3334377' (accession number O15393), a human transmembrane serine protease, a type -II membrane protein that is expressed strongly in small intestine. Paoloni-Giacobino et al., Genomics 44:309-320 (1997).

The nucleic acids and proteins related to SER6 are useful in potential therapeutic applications implicated in tumorigenesis and cancer, and/or other pathologies and disorders,

such as , for example, those involving blood clotting disorders, diseases of the mast cells involving tryptase, and intestinal disorders.

SER7 and SER8

Nucleic acids that are the reverse compliment of SER6 are shown below as SER7 (SEQ ID NO:16, Table 23, partial sequence) and SER8 (full sequence, SEQ ID NO: 17, Table 24).

Table 23: SER7 (12996895.0.1) IS A PARTIAL REVERSE COMPIMENT OF SER6

CAGAGTCTTCACAGCAGCCGGTGCTGTGCAGAGCAGCAGGCCAGCTGGTTAGGATTTTGTGAATCGCACCTCGCTCTCCA
TCTTGCTGTAAATCCAGGGAAGAACTTCTGTCACTTTGGTGTACACACCAGGTTTGTTTCTCTGGCCACAGCCTGTGCCC
CAGCTGGTGACACCTGCCAGGTACCAGCGGTTGTTCTGCTCACAGACAAGAGGCCCCCCGCTGTCTCCCTGGCAGGAGTC
TCTGCCCCCACGAAGGTCCCCAGCACACATCATCCTTGGGGTAAGGTAACTGTCATAGACCAAGTAGTCATTGCATTTCT
TGAAGTCGATGAGATTGACCTGCACCTCCCGGAGGAAGGGGGATGTCTTGTCATCTGTCTCCCTGGTCTTGCCAAAGCCT
GTGATCCAGCAGGTCTCATTGAGGCTAAAGGTCTGTCCATGCATGGGGAGGCAAGCAGGGTGGATGTGAGCGGACAGGGT
CAGGGGCTTGGACAGCCGCATGAGGGCGATGTCATAGTCGTCCTCCTCATCGGTGTAATTGCTGTTGATGATGATCTCGG
CAATGGAGGCTGCCTCAGGCAACTGGTGCAGGTTGCTGGTGCCCGCGTACACCTTCCAGCCCTCCAGGACCTTCTCCCGG
GTCACGAAGAAGCAGTGGGCGGCAGTGAGCACCCACTGGGCGTCAATGAGCGTGCCTCCACAGATGTGGGTGGTGCCGAA
GTGCAGACTCACTTGCCAAGGCCACTTGCTATCCGAGGCCAGCGCCCCTCCCACGATCCGCCCGGTCATGGCCCTCAGTC
CGCAGTGGGAACACTGGAGGGAGATATACCGCTGGGAAGGGCATTCAGACCTGTGGAGGCTTTCCTGGATGGTGGAGTTG
TATCTCAAGATTGAGAAGCTGTTGGCAAAATCCCTGTGGGCAACCTCGGTTGTCCGGTGAGCACTCTCGAAACCCAGCTG
CTGGCAGGTCTTCTCTGAGTAGGAGTCATTCCAGTTGCTGCTACAGATGGGAAGCCACTGATGGGAGGACCCAGAGTAGA
TTTTAAGCAGAGACTTGTCCCAGTCAAACCTCACGCAG (SEQ ID NO:16).

Table 24: SER8 IS A REVCOMP OF SER6

CAGAGTCTTCACAGCAGCCGGTGCTGTGCAGAGCAGCAGGCCAGCTGGTTAGGATTTTGTGAATCGCACCTCGCTCTCCA
TCTTGCTGTAAATCCAGGGAAGAACTTCTGTCACTTTGGTGTACACACCAGGTTTGTTTCTCTGGCCACAGCCTGTGCCC
CAGCTGGTGACACCTGCCAGGTACCAGCGGTTGTTCTGCTCACAGACAAGAGGCCCCCCGCTGTCTCCCTGGCAGGAGTC
TCTGCCCCCACGAAGGTCCCCAGCACACATCATCCTTGGGGTAAGGTAACTGTCATAGACCAAGTAGTCATTGCATTTCT
TGAAGTCGATGAGATTGACCTGCACCTCCCGGAGGAAGGGGGATGTCTTGTCATCTGTCTCCCTGGTCTTGCCAAAGCCT
GTGATCCAGCAGGTCTCATTGAGGCTAAAGGTCTGTCCATGCATGGGGAGGCAAGCAGGGTGGATGTGAGCGGACAGGGT
CAGGGGCTTGGACAGCCGCATGAGGGCGATGTCATAGTCGTCCTCCTCATCGGTGTAATTGCTGTTGATGATGATCTCGG
CAATGGAGGCTGCCTCAGGCAACTGGTGCAGGTTGCTGGTGCCCGCGTACACCTTCCAGCCCTCCAGGACCTTCTCCCGG
GTCACGAAGAAGCAGTGGGCGGCAGTGAGCACCCACTGGGCGTCAATGAGCGTGCCTCCACAGATGTGGGTGGTGCCGAA
GTGCAGACTCACTTGCCAAGGCCACTTGCTATCCGAGGCCAGCGCCCCTCCCACGATCCGCCCGGTCATGGCCCTCAGTC
CGCAGTGGGAACACTGGAGGGAGATATACCGCTGGGAAGGGCATTCAGACCTGTGGAGGCTTTCCTGGATGGTGGAGTTG
TATCTCAAGATTGAGAAGCTGTTGGCAAAATCCCTGTGGGCAACCTCGGTTGTCCGGTGAGCACTCTCGAAACCCAGCTG
CTGGCAGGTCTTCTCTGAGTAGGAGTCATTCCAGTTGCTGCTACAGATGGGAAGCCACTGATGGGAGGACCCAGAGTAGA
TTTTAAGCAGAGACTTGTCCCAGTCAAACCTCACGCAGCCCAGCTCGTCACTCTTCAGCTTGCAGTCCACCACCCCGTCA
CAGCGAACAGCGTGCTTGGGACAGCTCTCCCTCTGCTCCTTGTACCTGATCCCTGTGTGGCCCTGCCAGAACTGGAAGAG
GATGATGAGCGAAACCACCAGGGCAATGAGGAGGAGCACGCACCCGATGAGCGGTAGCTGCTTCTGGCCCTCCCGCCAGG
TGAACTTGGGCAGGCTCGTACCTGGGCTCTCCAT (SEQ ID NO:17)

Clone 12996895.0.1 (SER7) is predicted to encode a 342 aa secreted protein. SER7 has homology to several serine proteases. The highest homology is 201/344 (58%) with AAD37117 Transmembrane Serine Protease 2.

Quantitative expression analysis of clone 12996895.0.1, in various cells and tissues.

RTQ-PCR Panel 1 Description:

As shown in Table 26 below, this 96 well plate (2 control wells, 94 test samples) panel and its variants (Panel 1) are composed of RNA/cDNA isolated from various human cell lines that have been established from human malignant tissues (Tumors). These cell lines have been extensively characterized by investigators in both academia and the commercial sector regarding their tumorgenicity, metastatic potential, drug resistance, invasive potential and other cancer-related properties. They serve as suitable tools for pre-clinical evaluation of anti-cancer agents and promising therapeutic strategies. RNA from these various human cancer cell lines was isolated by and procured from the Developmental Therapeutic Branch (DTB) of the National Cancer Institute (USA). Basic information regarding their biological behavior, gene expression, and resistance to various cytotoxic agents are known in the art. In addition, RNA/cDNA was obtained from various human tissues derived from human autopsies performed on deceased elderly people or sudden death victims (accidents, etc.). These tissue were ascertained to be free of disease and were purchased from various high quality commercial sources such as Clontech, Inc., Research Genetics , and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electrophoresis using 28s and 18s ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the presence of low molecular weight RNAs indicative of degradation products. Samples are quality controlled for genomic DNA contamination by reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

RTQ-PCR Panel 2 Description-

As shown in Table 26 below, this 96 well (2 control wells, 94 test samples) panel and its variants (Panel 2) are composed of RNA/cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues procured are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" (NAT: normal adjacent tissue). The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery. In addition, RNA/cDNA was obtained from various human tissues derived from human autopsies performed on deceased elderly people or sudden death victims (accidents, etc.). These tissue were ascertained to be free of disease and were purchased from various high quality commercial sources such as Clontech, Inc., Research Genetics , and Invitrogen. RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electrophoresis using 28s and 18s ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the presence of low molecular weight RNAs indicative of degradation products. Samples are quality controlled for genomic DNA contamination by reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Methods:

The quantitative expression of various clones was assessed in about 41 normal and about 55 tumor samples by real time quantitative PCR (TaqMan®) performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. See Table 26.

First, 96 RNA samples were normalized to β-actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 48°C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using β-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2 min. at 50°C; 10 min. at 95°C; 15 sec. at 95°C/1 min. at 60°C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for ß-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their β-actin / GAPDH average CT values.

Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60° C, primer optimal Tm = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe T_m must be 10° C greater than primer T_m, amplicon size 75 bp to 100 bp. The probes and primers selected (see below, Table 25 were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by

HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

TaqMan oligo set Ag20 for the SER7 gene (*i.e.*, 12996895.0.1) include the forward (21 nt, probe (23 nt) and reverse (21 nt) oligomers. Sequences for the oligos are shown in Table 25.

Table 25: Taqman primers

Position	Primers	Sequences		Length
Start	Forward	5'-GTGGGAACACTGGAGGGAGAT-3'	SEQ ID NO:18	21
805	Probe	FAM-5'-AGGTCTGAATGCCCTTCCCAGCG-3'-TAMRA	SEQ ID NO:19	23
Reverse	Reverse	5'-CAACTCCACCATCCAGGAAAG-3'	SEQ ID NO:20	21

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (SEQX-specific and another gene-specific probe multiplexed with the SEQX probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/μl RNase inhibitor, and 0.25 U/μl reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute. The results are shown below in Table 26

Table 26: TaqMan Results

PANEL 1		Panel 2		
Tissue_Name/Run_Name	tm224f	Tissue_Name/Run_Name	2tm1317f	2tm1701f
Endothelial cells	0.1	Normal Colon GENPAK 061003	1.4	1.2
Endothelial cells (treated)	0.0	83219 CC Well to Mod Diff (ODO3866)	7.3	2.9
Pancreas	6.8	83220 CC NAT (ODO3866)	0.7	0.2
Pancreatic ca. CAPAN 2	0.2	83221 CC Gr.2 rectosigmoid (ODO3868)	10.6	4.9
Adipose	6.0	83222 CC NAT (ODO3868)	0.5	0.3
Adrenal gland	0.1	83235 CC Mod Diff (ODO3920)	43.5	40.6
Thyroid	26.2	83236 CC NAT (ODO3920)	4.2	2.7
Salivary gland	15.8	83237 CC Gr.2 ascend colon (ODO3921)	2.0	2.4
Pituitary gland	0.4	83238 CC NAT (ODO3921)	0.8	0.3
Brain (fetal)	0.8	83241 CC from Partial Hepatectomy (ODO4309)	5.7	3.9
Brain (whole)	0.7	83242 Liver NAT (ODO4309)	12.1	8.8
Brain (amygdala)	0.8	87472 Colon mets to lung (OD04451-01)	17.2	9.9
Brain (cerebellum)	0.1	87473 Lung NAT (OD04451-02)	4.2	5.0
Brain (hippocampus)	3.4	Normal Prostate Clontech A+ 6546-1	0.3	0.4
Brain (substantia nigra)	0.2	84140 Prostate Cancer (OD04410)	5.0	3.3
Brain (thalamus)	0.1	84141 Prostate NAT (OD04410)	7.3	3.3
Brain (hypothalamus)	0.8	87073 Prostate Cancer (OD04720-01)	100.0	100.0
Spinal cord	0.1	87074 Prostate NAT (OD04720-02)	9.3	11.5
CNS ca. (glio/astro) U87-MG	0.0	Normal Lung GENPAK 061010	1.2	2.4
CNS ca. (glio/astro) U-118-MG	0.1	83239 Lung Met to Muscle (ODO4286)	0.2	0.4
CNS ca. (astro) SW1783	0.1	83240 Muscle NAT (ODO4286)	3.9	2.6
CNS ca.* (neuro; met) SK-N-AS	0.0	84136 Lung Malignant Cancer (OD03126)	11.9	13.7
CNS ca. (astro) SF-539	0.0	84137 Lung NAT (OD03126)	4.1	1.4
CNS ca. (astro) SNB-75	0.0	84871 Lung Cancer (OD04404)	5.2	2.8
CNS ca. (glio) SNB-19	0.2	84872 Lung NAT (OD04404)	1.4	3.7
CNS ca. (glio) U251	0.0	84875 Lung Cancer (OD04565)	10.2	11.3
CNS ca. (glio) SF-295	0.1	85950 Lung Cancer (OD04237-01)	32.3	18.7
Heart	0.1	85970 Lung NAT (OD04237-02)	13.9	6.4
Skeletal muscle	0.0	83255 Ocular Mel Met to Liver (ODO4310)	1.1	0.9
Bone marrow	0.6	83256 Liver NAT (ODO4310)	8.1	5.5
Thymus	2.3	84139 Melanoma Mets to Lung (OD04321)	2.4	2.3
Spleen	1.9	84138 Lung NAT (OD04321)	4.2	1.7
Lymph node	8.0	Normal Kidney GENPAK 061008	2.7	1.3
Colon (ascending)	1.8	83786 Kidney Ca, Nuclear grade 2 (OD04338)	2.1	1.1
Stomach	1.6	83787 Kidney NAT (OD04338)	4.6	3.8
Small intestine	0.4	83788 Kidney Ca Nuclear grade ½ (OD04339)	12.9	13.2
Colon ca. SW480	0.1	83789 Kidney NAT (OD04339)	5.2	4.8
Colon ca.* (SW480 met)SW620	0.0	83790 Kidney Ca, Clear cell type (OD04340)	2.7	4.8
Colon ca. HT29	4.1	83791 Kidney NAT (OD04340)	4.0	2.8

Colon ca. HCT-116	0.0	83792 Kidney Ca, Nuclear grade 3 (OD04348)	2.2	1.3
Colon ca. CaCo-2	22.4	83793 Kidney NAT (OD04348)	7.4	3.2
Colon ca. HCT-15	9.2	87474 Kidney Cancer (OD04622-01)	5.2	4.9
Colon ca. HCC-2998	0.4	87475 Kidney NAT (OD04622-03)	3.3	2.4
Gastric ca.* (liver met) NCI-N87	13.9	85973 Kidney Cancer (OD04450-01)	0.2	0.7
Bladder	0.5	85974 Kidney NAT (OD04450-03)	6.4	6.3
Trachea	1.2	Kidney Cancer Clontech 8120607	1.8	3.0
Kidney	1.7	Kidney NAT Clontech 8120608	1.9	1.3
Kidney (fetal)	1.9	Kidney Cancer Clontech 8120613	0.0	0.4
Renal ca. 786-0	0.0	Kidney NAT Clontech 8120614	1.6	1.2
Renal ca. A498	0.1	Kidney Cancer Clontech 9010320	1.0	1.0
Renal ca. RXF 393	0.2	Kidney NAT Clontech 9010321	1.0	1.0
Renal ca. ACHN	0.5	Normal Uterus GENPAK 061018	1.8	0.8
Renal ca. UO-31	0.2	Uterus Cancer GENPAK 064011	9.5	4.5
Renal ca. TK-10	0.0	Normal Thyroid Clontech A+ 6570-1**	0.2	0.3
Liver	1.0	Thyroid Cancer GENPAK 064010	0.6	0.9
Liver (fetal)	1.2	Thyroid Cancer INVITROGEN A302152	5.0	12.1
Liver ca. (hepatoblast) HepG2	0.0	Thyroid NAT INVITROGEN A302153	25.7	6.5
Lung	0.9	Normal Breast GENPAK 061019	14.3	15.7
Lung (fetal)	12.2	84877 Breast Cancer (OD04566)	14.6	12.7
Lung ca. (small cell) LX-1	0.5	85975 Breast Cancer (OD04590-01)	36.6	41.8
Lung ca. (small cell) NCI-H69	3.2	85976 Breast Cancer Mets (OD04590-03)	38.2	34.4
Lung ca. (s.cell var.) SHP-77	0.0	87070 Breast Cancer Metastasis (OD04655-05)	85.3	51.8
Lung ca. (large cell)NCI-H460	0.0	GENPAK Breast Cancer 064006	50.4	37.9
Lung ca. (non-sm. cell) A549	0.5	Breast Cancer Clontech 9100266	13.5	12.9
Lung ca. (non-s.cell) NCI-H23	0.2	Breast NAT Clontech 9100265	11.0	15.7
Lung ca (non-s.cell) HOP-62	0.0	Breast Cancer INVITROGEN A209073	19.0	10.4
Lung ca. (non-s.cl) NCI-H522	0.2	Breast NAT INVITROGEN A2090734	12.0	12.0
Lung ca. (squam.) SW 900	38.2	Normal Liver GENPAK 061009	2.1	2.6
Lung ca. (squam.) NCI-H596	1.0	Liver Cancer GENPAK 064003	0.8	0.3
Mammary gland	28.1	Liver Cancer Research Genetics RNA 1025	2.8	2.6
Breast ca.* (pl. effusion) MCF-7	92.7	Liver Cancer Research Genetics RNA 1026	0.5	0.1
Breast ca.* (pl.ef) MDA-MB-231	0.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	1.2	1.6
Breast ca.* (pl. effusion) T47D	100.0	Paired Liver Tissue Research Genetics RNA 6004-N	3.3	2.5
Breast ca. BT-549	0.0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.5	0.4
Breast ca. MDA-N	0.4	Paired Liver Tissue Research Genetics RNA 6005-N	1.2	1.1
Ovary	0.2	Normal Bladder GENPAK 061001	7.9	2.7
Ovarian ca. OVCAR-3	2.8	Bladder Cancer Research Genetics RNA 1023	11.0	9.0
Ovarian ca. OVCAR-4	2.2	Bladder Cancer INVITROGEN A302173	2.8	6.7
Ovarian ca. OVCAR-5	2.7	87071 Bladder Cancer (OD04718-01)	11.7	10.7
Ovarian ca. OVCAR-8	0.2	87072 Bladder Normal Adjacent (OD04718-03)	24.3	6.8
Ovarian ca. IGROV-1	0.2	Normal Ovary Res. Gen.	0.2	0.2

Ovarian ca.* (ascites) SK-OV-3	0.1	Ovarian Cancer GENPAK 064008	12.1	6.9
Uterus	1.3	87492 Ovary Cancer (OD04768-07)	12.2	10.1
Placenta	83.5	87493 Ovary NAT (OD04768-08)	2.2	1.0
Prostate	3.1	Normal Stomach GENPAK 061017	0.2	0.2
Prostate ca.* (bone met)PC-3	0.0	NAT Stomach Clontech 9060359	0.3	0.5
Testis	1.6	Gastric Cancer Clontech 9060395	0.5	0.2
Melanoma Hs688(A).T	0.0	NAT Stomach Clontech 9060394	0.3	1.1
Melanoma* (met) Hs688(B).T	0.1	Gastric Cancer Clontech 9060397	0.3	0.4
Melanoma UACC-62	0.0	NAT Stomach Clontech 9060396	0.4	0.6
Melanoma M14	4.3	Gastric Cancer GENPAK 064005	1.0	1.6
Melanoma LOX IMVI	1.0			
Melanoma* (met) SK-MEL-5	0.1			
Melanoma SK-MEL-28	0.5			

In Table 26 the following abbreviations are used:
ca. = carcinoma,
* = established from metastasis,
met = metastasis,
s cell var= small cell variant,
non-s = non-sm =non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,
glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

The results presented in Table 26 show that there is high expression of clone 12996895.0.1 in breast cancer, lung cancer, thyroid cancer, prostate cancer, colon cancer, placenta, and moderate to low expression in salivary gland, normal prostate, normal colon and normal breast tissue. This result suggests that the gene or its gene product may potentially be useful use in therapeutic approaches for cancers and related hypo- and hyperproliferative cell disorders. For example, SER7 and SER8 may be involved in autocrine stimulation of tumor growth, angiogenesis and metastatic progression. Also these SERX proteins may have a role in stimulating tumor cell matrix degradation and tumor cell migration (*e.g.*, invasion). Based on the above data, targeting of SER7 and SER8 with a monoclonal antibody may have an inhibitory effect on tumor growth and progression. See, e.g., Kazama, et al, J. Biol. Chem. 270:66-72 (1995) and Noel et al., Invasion Metastasis 17:221-239 (1997).

SERX Nucleic Acids

The nucleic acids of the invention include those that encode a SERX polypeptide or protein. As used herein, the terms polypeptide and protein are interchangeable.

In some embodiments, a SERX nucleic acid encodes a mature SERX polypeptide. As used herein, a "mature" form of a polypeptide or protein described herein relates to the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an open reading frame described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

Among the SERX nucleic acids is the nucleic acid whose sequence is provided SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, or a fragment thereof, any of whose bases may be changed from the corresponding bases shown in SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, while still encoding a protein that maintains at least one of its SERX-like activities and physiological functions (*i.e.*, modulating angiogenesis, neuronal development). The invention further includes the complement of the nucleic acid sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, including fragments, derivatives, analogs and homologs thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

One aspect of the invention pertains to isolated nucleic acid molecules that encode SERX proteins or biologically active portions thereof. Also included are nucleic acid fragments sufficient for use as hybridization probes to identify SERX-encoding nucleic acids (*e.g.*, SERX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of SERX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated SERX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, or a complement of any of

this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, as a hybridization probe, SERX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to SERX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at lease 6 contiguous nucleotides SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 is one that is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 that it can hydrogen bond with little or no mismatches to the

nucleotide sequence shown in SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotide units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of SERX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the

alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of a SERX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a SERX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human SERX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2, 4, 6 or 8, as well as a polypeptide having SERX activity. Biological activities of the SERX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human SERX polypeptide.

The nucleotide sequence determined from the cloning of the human SERX gene allows for the generation of probes and primers designed for use in identifying and/or cloning SERX homologues in other cell types, *e.g.*, from other tissues, as well as SERX homologues from other mammals. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350

or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17; or an anti-sense strand nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17; or of a naturally occurring mutant of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17.

Probes based on the human SERX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a SERX protein, such as by measuring a level of a SERX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting SERX mRNA levels or determining whether a genomic SERX gene has been mutated or deleted.

A "polypeptide having a biologically active portion of SERX" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of SERX" can be prepared by isolating a portion of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 that encodes a polypeptide having a SERX biological activity (biological activities of the SERX proteins are described below), expressing the encoded portion of SERX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of SERX. For example, a nucleic acid fragment encoding a biologically active portion of SERX can optionally include an ATP-binding domain. In another embodiment, a nucleic acid fragment encoding a biologically active portion of SERX includes one or more regions.

SERX Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 due to the degeneracy of the genetic code. These nucleic acids thus encode the same SERX protein as that encoded by the nucleotide sequence shown in SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 or 7 *e.g.*, the polypeptide of SEQ ID NO: 2, 6, 9, 11, or 15. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO: 2

In addition to the human SERX nucleotide sequence shown in SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of SERX may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the SERX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a SERX protein, preferably a mammalian SERX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the SERX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in SERX that are the result of natural allelic variation and that do not alter the functional activity of SERX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding SERX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the SERX cDNAs of the invention can be isolated based on their homology to the human SERX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a soluble human SERX cDNA can be isolated based on its homology to human membrane-bound SERX. Likewise, a membrane-bound human SERX cDNA can be isolated based on its homology to soluble human SERX.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding SERX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high

stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65 °C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50 °C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17,

or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40 °C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50 °C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA 78*: 6789-6792.

Conservative mutations

In addition to naturally-occurring allelic variants of the SERX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, thereby leading to changes in the amino acid sequence of the encoded SERX protein, without altering the functional ability of the SERX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of SERX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the SERX proteins of the present invention, are predicted to be particularly unamenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding SERX proteins that contain changes in amino acid residues that are not essential for activity. Such SERX proteins differ in amino acid sequence from SEQ ID NO: 2, 6, 9, 11, or 15, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of SEQ ID NO: 2, 6, 9, 11, or 15. Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to SEQ ID NO: 2, 6, 9, 11, or 15, more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO: 2, 6, 9, 11, or 15.

An isolated nucleic acid molecule encoding a SERX protein homologous to the protein of can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into the nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in SERX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a SERX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for SERX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant SERX protein can be assayed for (1) the ability to form protein:protein interactions with other SERX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant SERX protein and a SERX receptor; (3) the ability of a mutant SERX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind SERX protein; or (5) the ability to specifically bind an anti-SERX protein antibody.

Antisense SERX Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire SERX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a SERX protein of SEQ ID NO: 2, 6, 9, 11, or 15 or antisense nucleic acids complementary to a SERX nucleic acid sequence of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding SERX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the protein coding region of human SERX corresponds to SEQ ID NO: 2, 6, 9, 11, or 15). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding SERX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding SERX disclosed herein (*e.g.*, SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of SERX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of SERX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of SERX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense

nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a SERX protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix: An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface,

e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

SERX Ribozymes and PNA moieties

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as a mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave SERX mRNA transcripts to thereby inhibit translation of SERX mRNA. A ribozyme having specificity for a SERX-encoding nucleic acid can be designed based upon the nucleotide sequence of a SERX DNA disclosed herein (*i.e.*, SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a SERX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, SERX mRNA can be used to select a catalytic RNA having a specific

ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, SERX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the SERX (*e.g.*, the SERX promoter and/or enhancers) to form triple helical structures that prevent transcription of the SERX gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of SERX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of SERX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of SERX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of SERX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of SERX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA

chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl) amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

SERX Polypeptides

A SERX polypeptide of the invention includes the SERX-like protein whose sequence is provided in SEQ ID NO: 2, 6, 9, 11, or 15. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in SEQ ID NO: 2, 6, 9, 11, or 15 while still encoding a protein that maintains its SERX-like activities and physiological functions, or a functional fragment thereof. In some embodiments, up to 20% or more of the residues may be so changed in the mutant or variant protein. In some embodiments, the SERX polypeptide according to the invention is a mature polypeptide.

In general, a SERX -like variant that preserves SERX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated SERX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-SERX antibodies. In one embodiment, native SERX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, SERX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a SERX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the SERX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of SERX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of SERX protein having less than about 30% (by dry weight) of non-SERX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-SERX protein, still more preferably less than about 10% of non-SERX protein, and most preferably less than about 5% non-SERX protein. When the SERX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of SERX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the

language "substantially free of chemical precursors or other chemicals" includes preparations of SERX protein having less than about 30% (by dry weight) of chemical precursors or non-SERX chemicals, more preferably less than about 20% chemical precursors or non-SERX chemicals, still more preferably less than about 10% chemical precursors or non-SERX chemicals, and most preferably less than about 5% chemical precursors or non-SERX chemicals.

Biologically active portions of a SERX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the SERX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO: 2, 6, 9, 11, or 15 that include fewer amino acids than the full length SERX proteins, and exhibit at least one activity of a SERX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the SERX protein. A biologically active portion of a SERX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a SERX protein of the present invention may contain at least one of the above-identified domains conserved between the SERX proteins, *e.g.* TSR modules. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native SERX protein.

In an embodiment, the SERX protein has an amino acid sequence shown in SEQ ID NO: 2, 6, 9, 11, or 15. In other embodiments, the SERX protein is substantially homologous to SEQ ID NO: 2, 6, 9, 11, or 15 and retains the functional activity of the protein of SEQ ID NO: 2, 6, 9, 11, or 15 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the SERX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO: 2, 6, 9, 11, or 15 and retains the functional activity of the SERX proteins of SEQ ID NO: 2, 6, 9, 11, or 15.

Determining homology between two or more sequence

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide

positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, *Needleman and Wunsch* 1970 *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

Chimeric and fusion proteins

The invention also provides SERX chimeric or fusion proteins. As used herein, a SERX "chimeric protein" or "fusion protein" comprises a SERX polypeptide operatively linked to a non-SERX polypeptide. An "SERX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to SERX, whereas a "non-SERX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the SERX protein, *e.g.*, a protein that is different from the SERX protein and that is derived from the same or a different organism. Within a SERX fusion protein the SERX polypeptide can correspond to all or a portion of a SERX protein. In one embodiment, a SERX fusion protein comprises at least one biologically active portion of a SERX protein. In another embodiment, a SERX fusion protein comprises at least two biologically active portions of a SERX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the SERX polypeptide and the non-SERX polypeptide are fused in-frame to each other. The non-SERX polypeptide can be fused to the N-terminus or C-terminus of the SERX polypeptide.

For example, in one embodiment a SERX fusion protein comprises a SERX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate SERX activity (such assays are described in detail below).

In another embodiment, the fusion protein is a GST-SERX fusion protein in which the SERX sequences are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant SERX.

In another embodiment, the fusion protein is a SERX-immunoglobulin fusion protein in which the SERX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The SERX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a SERX ligand and a SERX protein on the surface of a cell, to thereby suppress SERX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated SERX ligand of the invention is the SERX receptor. The SERX-immunoglobulin fusion proteins can be used to affect the bioavailability of a SERX cognate ligand. Inhibition of the SERX ligand/SERX interaction may be useful

therapeutically for both the treatment of proliferative and differentiative disorders, *e,g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the SERX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-SERX antibodies in a subject, to purify SERX ligands, and in screening assays to identify molecules that inhibit the interaction of SERX with a SERX ligand.

A SERX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A SERX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the SERX protein.

SERX agonists and antagonists

The present invention also pertains to variants of the SERX proteins that function as either SERX agonists (mimetics) or as SERX antagonists. Variants of the SERX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the SERX protein. An agonist of the SERX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the SERX protein. An antagonist of the SERX protein can inhibit one or more of the activities of the naturally occurring form of the SERX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the SERX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring

form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the SERX proteins.

Variants of the SERX protein that function as either SERX agonists (mimetics) or as SERX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the SERX protein for SERX protein agonist or antagonist activity. In one embodiment, a variegated library of SERX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of SERX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential SERX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of SERX sequences therein. There are a variety of methods which can be used to produce libraries of potential SERX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential SERX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

Polypeptide libraries

In addition, libraries of fragments of the SERX protein coding sequence can be used to generate a variegated population of SERX fragments for screening and subsequent selection of variants of a SERX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a SERX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the SERX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of SERX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify SERX variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

SERX Antibodies

Also included in the invention are antibodies to SERX proteins, or fragments of SERX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated SERX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the

amino acid sequence of the full length protein and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of SERX-related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human SERX-related protein sequence will indicate which regions of a SERX-related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native

protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by

immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human

variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab}

fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the

large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has

provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc R), such as Fc RI (CD64), Fc RII (CD32) and Fc RIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives

(such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

SERX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a SERX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such

other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, SERX proteins, mutant forms of SERX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of SERX proteins in prokaryotic or eukaryotic cells. For example, SERX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the SERX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30:

933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, SERX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to SERX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, SERX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as human, Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.*

(MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding SERX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) SERX protein. Accordingly, the invention further provides methods for producing SERX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding SERX protein has been introduced) in a suitable medium such that SERX protein is produced. In another embodiment, the method further comprises isolating SERX protein from the medium or the host cell.

Transgenic SERX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which SERX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous SERX sequences have been introduced into their genome or homologous recombinant animals in which endogenous SERX sequences have been altered. Such animals are useful for studying the function and/or activity of SERX protein and for identifying and/or evaluating modulators of SERX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals

include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous SERX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing SERX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. Sequences including SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human SERX gene, such as a mouse SERX gene, can be isolated based on hybridization to the human SERX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the SERX transgene to direct expression of SERX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the SERX transgene in its genome and/or expression of SERX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding SERX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a SERX gene into which a deletion, addition or substitution has been

introduced to thereby alter, *e.g.*, functionally disrupt, the SERX gene. The SERX gene can be a human gene (*e.g.*, the DNA of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17), but more preferably, is a non-human homologue of a human SERX gene. For example, a mouse homologue of human SERX gene of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17can be used to construct a homologous recombination vector suitable for altering an endogenous SERX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous SERX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous SERX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous SERX protein). In the homologous recombination vector, the altered portion of the SERX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the SERX gene to allow for homologous recombination to occur between the exogenous SERX gene carried by the vector and an endogenous SERX gene in an embryonic stem cell. The additional flanking SERX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced SERX gene has homologously-recombined with the endogenous SERX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are

described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae. See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The SERX nucleic acid molecules, SERX proteins, and anti-SERX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like,

compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al ., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon et al., J. National Cancer Inst., 81(19): 1484 (1989).

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates,

and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a SERX protein or anti-SERX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of

the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable,

biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed.

(Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York. If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, *e.g.*, Marasco *et al.*, 1993 *Proc. Natl. Acad. Sci. USA*, 90: 7889-7893. The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended. The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and ethyl-L-glutamate, non-degradable ethylene-vinyl acetate,

degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express SERX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect SERX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a SERX gene, and to modulate SERX activity, as described further, below. In addition, the SERX proteins can be used to screen drugs or compounds that modulate the SERX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of SERX protein or production of SERX protein forms that have decreased or aberrant activity compared to SERX wild-type protein . In addition, the anti-SERX antibodies of the invention can be used to detect and isolate SERX proteins and modulate SERX activity. For example, SERX activity includes growth and differentiation, antibody production, and tumor growth.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to SERX proteins or have a stimulatory or inhibitory effect on, *e.g.*, SERX protein expression or SERX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a SERX protein or polypeptide or biologically-active portion thereof. The test compounds of the

invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of SERX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a SERX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the SERX protein can be

accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the SERX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of SERX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds SERX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a SERX protein, wherein determining the ability of the test compound to interact with a SERX protein comprises determining the ability of the test compound to preferentially bind to SERX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of SERX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the SERX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of SERX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the SERX protein to bind to or interact with a SERX target molecule. As used herein, a "target molecule" is a molecule with which a SERX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a SERX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A SERX target molecule can be a non-SERX molecule or a SERX protein or polypeptide of the invention In one embodiment, a SERX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound SERX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that

has catalytic activity or a protein that facilitates the association of downstream signaling molecules with SERX.

Determining the ability of the SERX protein to bind to or interact with a SERX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the SERX protein to bind to or interact with a SERX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a SERX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a SERX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the SERX protein or biologically-active portion thereof. Binding of the test compound to the SERX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the SERX protein or biologically-active portion thereof with a known compound which binds SERX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a SERX protein, wherein determining the ability of the test compound to interact with a SERX protein comprises determining the ability of the test compound to preferentially bind to SERX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting SERX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the SERX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of SERX can be accomplished, for example, by determining the ability of the SERX protein to bind to a SERX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of SERX protein can be accomplished by determining the

ability of the SERX protein further modulate a SERX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described above.

In yet another embodiment, the cell-free assay comprises contacting the SERX protein or biologically-active portion thereof with a known compound which binds SERX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a SERX protein, wherein determining the ability of the test compound to interact with a SERX protein comprises determining the ability of the SERX protein to preferentially bind to or modulate the activity of a SERX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of SERX protein. In the case of cell-free assays comprising the membrane-bound form of SERX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of SERX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either SERX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to SERX protein, or interaction of SERX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-SERX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test

compound and either the non-adsorbed target protein or SERX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of SERX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the SERX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated SERX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with SERX protein or target molecules, but which do not interfere with binding of the SERX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or SERX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the SERX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the SERX protein or target molecule.

In another embodiment, modulators of SERX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of SERX mRNA or protein in the cell is determined. The level of expression of SERX mRNA or protein in the presence of the candidate compound is compared to the level of expression of SERX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of SERX mRNA or protein expression based upon this comparison. For example, when expression of SERX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of SERX mRNA or protein expression. Alternatively, when expression of SERX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate

compound is identified as an inhibitor of SERX mRNA or protein expression. The level of SERX mRNA or protein expression in the cells can be determined by methods described herein for detecting SERX mRNA or protein.

In yet another aspect of the invention, the SERX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with SERX ("SERX-binding proteins" or "SERX-bp") and modulate SERX activity. Such SERX-binding proteins are also likely to be involved in the propagation of signals by the SERX proteins as, for example, upstream or downstream elements of the SERX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for SERX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a SERX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with SERX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) identify an individual

from a minute biological sample (tissue typing); and (*ii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Tissue Typing

The SERX sequences of the invention can be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the SERX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The SERX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in

SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining SERX protein and/or nucleic acid expression as well as SERX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant SERX expression or activity. Disorders associated with aberrant SERX expression of activity include, for example, neurodegenerative, cell proliferative, angiogenic, hematopoietic, immunological, inflammatory, and tumor-related disorders and/or pathologies.

The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with SERX protein, nucleic acid expression or activity. For example, mutations in a SERX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with SERX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining SERX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of SERX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of SERX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting SERX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes SERX protein such that the presence of SERX is detected in the biological sample. An agent for detecting SERX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to SERX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length SERX nucleic acid, such as the nucleic acid of SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to SERX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

One agent for detecting SERX protein is an antibody capable of binding to SERX protein, preferably an antibody with a detectable label. Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds.

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive

materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect SERX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of SERX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of SERX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of SERX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of SERX protein include introducing into a subject a labeled anti-SERX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In one embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting SERX protein, mRNA, or genomic DNA, such that the presence of SERX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of SERX protein, mRNA or genomic DNA in the control sample with the presence of SERX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of SERX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting SERX protein or mRNA in a biological sample; means for determining the amount of SERX in the sample; and means for comparing the amount of SERX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect SERX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant SERX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with SERX protein, nucleic acid expression or activity. Such disorders include for example, neurodegenerative, cell proliferative, angiogenic, hematopoietic, immunological, inflammatory, and tumor-related disorders and/or pathologies.

Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant SERX expression or activity in which a test sample is obtained from a subject and SERX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of SERX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant SERX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant SERX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant SERX expression or activity in which a test sample is obtained and SERX protein or nucleic acid is detected (*e.g.*, wherein the presence of SERX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant SERX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a SERX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a SERX-protein, or the misexpression of the SERX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from a SERX gene; (*ii*) an addition of one or more nucleotides to a SERX gene; (*iii*) a substitution of one or more nucleotides of a SERX gene, (*iv*) a chromosomal rearrangement of a SERX gene; (*v*) an alteration in the level of a messenger RNA transcript of a SERX gene, (*vi*) aberrant modification of a SERX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a SERX gene, (*viii*) a non-wild-type level of a SERX protein, (*ix*) allelic loss of a SERX gene, and (*x*) inappropriate post-translational modification of a SERX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a SERX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*,

Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the SERX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a SERX gene under conditions such that hybridization and amplification of the SERX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a SERX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in SERX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in SERX can be identified in two dimensional arrays containing light-generated

DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the SERX gene and detect mutations by comparing the sequence of the sample SERX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the SERX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type SERX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine

the site of mutation. *See, e.g.*, Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in SERX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.*, Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a SERX sequence, *e.g.*, a wild-type SERX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in SERX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.*, Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control SERX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.*, Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.*, Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely

denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a SERX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which SERX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on SERX activity (*e.g.*, SERX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.*, neurodegenerative, cell proliferative, angiogenic, hematopoietic, immunological, inflammatory, and tumor-related disorders and/or pathologies). In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of SERX protein, expression of SERX nucleic acid, or mutation content of SERX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main

clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of SERX protein, expression of SERX nucleic acid, or mutation content of SERX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a SERX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of SERX (*e.g.*, the ability to modulate aberrant cell proliferation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an

agent determined by a screening assay as described herein to increase SERX gene expression, protein levels, or upregulate SERX activity, can be monitored in clinical trails of subjects exhibiting decreased SERX gene expression, protein levels, or downregulated SERX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease SERX gene expression, protein levels, or downregulate SERX activity, can be monitored in clinical trails of subjects exhibiting increased SERX gene expression, protein levels, or upregulated SERX activity. In such clinical trials, the expression or activity of SERX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including SERX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates SERX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of SERX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of SERX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a SERX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the SERX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the SERX

protein, mRNA, or genomic DNA in the pre-administration sample with the SERX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of SERX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of SERX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant SERX expression or activity. For example, disorders associated with aberrant SER1-4 expression of activity include a cell cycl or hyper- or hypo-proliferative disorder, *e.g.*, breast or testicular cancer, a cardiovascular defect, *e.g.*, DGS or VCFS and fertility or reproductive related disorders, autoimmune diseases, asthma, emphysema, scleroderma and developmental disorders.

Whereas, disorders associated with aberrant SER5-8 expression include, for example,various blood clotting hematopoietic, and tumor-related (e.g., osteosarcoma, hepatoma) disorders and/or pathologies.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors,

agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant SERX expression or activity, by administering to the subject an agent that modulates SERX expression or at least one SERX activity. Subjects at risk for a disease that is caused or contributed to by aberrant SERX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the SERX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of SERX aberrancy, for example, a SERX agonist or SERX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating SERX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of SERX protein activity associated with the cell. An agent that modulates SERX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a SERX protein, a peptide, a SERX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more SERX protein activity. Examples of such stimulatory agents include active SERX protein and a nucleic acid molecule encoding SERX that has been introduced into the cell. In another embodiment, the agent inhibits one or more SERX protein activity. Examples of such inhibitory agents include antisense SERX nucleic acid molecules and anti-SERX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a SERX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) SERX expression or activity. In another embodiment, the method involves administering a SERX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant SERX expression or activity.

Stimulation of SERX activity is desirable in situations in which SERX is abnormally downregulated and/or in which increased SERX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated).

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the

antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo*

testing, any of the animal model system known in the art may be used prior to administration to human subjects.

OTHER EMBODIMENTS

While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) a mature form of the amino acid sequence given by SEQ ID NO: 2, 6, 9, 11, or 15;

b) a variant of a mature form of the amino acid sequence given by SEQ ID NO: 2, 6, 9, 11, or 15, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed;

c) the amino acid sequence given by SEQ ID NO: 2, 6, 9, 11, or 15;

d) a variant of the amino acid sequence given by SEQ ID NO: 2, 6, 9, 11, or 15 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and

e) a fragment of any of a) through d).

2. The polypeptide of claim 1 that is a naturally occurring allelic variant of the sequence given by SEQ ID NO: 2, 6, 9, 11, or 15.

3. The polypeptide of claim 2, wherein the variant is the translation of a single nucleotide polymorphism.

4. The polypeptide of claim 1 that is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution.

5. An isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of:

a) a mature form of the amino acid sequence given SEQ ID NO: 2, 6, 9, 11, or 15;

b) a variant of a mature form of the amino acid sequence given by SEQ ID NO: 2, 6, 9, 11, or 15 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed;

c) the amino acid sequence given by SEQ ID NO: 2, 6, 9, 11, or 15;

d) a variant of the amino acid sequence given by SEQ ID NO: 2, 6, 9, 11, or 15, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed;

e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence given by SEQ ID NO: 2, 6, 9, 11, or 15 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and

f) the complement of any of said nucleic acid molecules.

6. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

7. The nucleic acid molecule of claim 5 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

8. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises a single nucleotide polymorphism encoding said variant polypeptide.

9. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of

a) the nucleotide sequence given by SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17;

b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence given by SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17is changed from that given by the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed;

c) a nucleic acid fragment of the sequence given by SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17; and

d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence given by SEQ ID NO: 1, 5, 7, 8, 10, 14, 16, or 17 is changed from that given by the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

10. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence given by SEQ ID NO1, 5, 7, 8, 10, 14, 16, or 17, or a complement of said nucleotide sequence.

11. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that given by the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides in the chosen coding sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

12. A vector comprising the nucleic acid molecule of claim 11.

13. The vector of claim 12, further comprising a promoter operably linked to said nucleic acid molecule.

14. A cell comprising the vector of claim 12.

15. An antibody that binds immunospecifically to the polypeptide of claim 1.

16. The antibody of claim 15, wherein said antibody is a monoclonal antibody.

17. The antibody of claim 15, wherein the antibody is a humanized antibody.

18. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and

(c) determining the presence or amount of antibody bound to said polypeptide,

thereby determining the presence or amount of polypeptide in said sample.

19. A method for determining the presence or amount of the nucleic acid molecule of claim 5 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of said probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

20. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

(a) introducing said polypeptide to said agent; and

(b) determining whether said agent binds to said polypeptide.

21. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

(a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;

(b) contacting the cell with a composition comprising a candidate substance; and

(c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

22. A method for modulating the activity of the polypeptide of claim 1, the method comprising introducing a cell sample expressing the polypeptide of said claim with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

23. A method of treating or preventing a pathology associated with the polypeptide of claim 1, said method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent said pathology in said subject.

24. The method of claim 23, wherein said subject is a human.

25. A method of treating or preventing a pathology associated with the polypeptide of claim 1, said method comprising administering to a subject in which such treatment or prevention is desired a SERX nucleic acid in an amount sufficient to treat or prevent said pathology in said subject.

26. The method of claim 25, wherein said subject is a human.

27. A method of treating or preventing a pathology associated with the polypeptide of claim 1, said method comprising administering to a subject in which such treatment or prevention is desired a SERX antibody in an amount sufficient to treat or prevent said pathology in said subject.

28. The method of claim 15, wherein the subject is a human.

29. A pharmaceutical composition comprising the polypeptide of claim 1 and a pharmaceutically acceptable carrier.

30. A pharmaceutical composition comprising the nucleic acid molecule of claim 5 and a pharmaceutically acceptable carrier.

31. A pharmaceutical composition comprising the antibody of claim 15 and a pharmaceutically acceptable carrier.

32. A kit comprising in one or more containers, the pharmaceutical composition of claim 29.

33. A kit comprising in one or more containers, the pharmaceutical composition of claim 30.

34. A kit comprising in one or more containers, the pharmaceutical composition of claim 31.

35. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein said therapeutic is the polypeptide of claim 1.

36. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein said therapeutic is a SERX nucleic acid.

37. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein said therapeutic is a SERX antibody.

38. A method for screening for a modulator of activity or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;

b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and

c) comparing the activity of said protein in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the polypeptide of claim 1.

39. The method of claim 38, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

40. A method for determining the presence of or predisposition to a disease associated with altered levels of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and

b) comparing the amount of said polypeptide in the sample of step (a) to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

41. A method for determining the presence of or predisposition to a disease associated with altered levels of the nucleic acid molecule of claim 5 in a first mammalian subject, the method comprising:

a) measuring the amount of the nucleic acid in a sample from the first mammalian subject; and